

Victor Maia Senna Delgado

Eficiência das Escolas Públicas Estaduais de Minas Gerais:  
considerações acerca da qualidade a partir da análise dos  
dados do SICA e do SIMAVE

Belo Horizonte, MG  
UFMG/CEDEPLAR  
2007

Victor Maia Senna Delgado

Eficiência das Escolas Públicas Estaduais de Minas Gerais:  
considerações acerca da qualidade a partir da análise dos  
dados do SICA e do SIMAVE

Dissertação apresentada ao Curso de  
Mestrado em Economia do Centro de  
Desenvolvimento e Planejamento Regional  
da Faculdade de Ciências Econômicas da  
Universidade Federal de Minas Gerais,  
como requisito parcial à obtenção do título  
de mestre em Economia.

Orientadora: Profa<sup>a</sup> Ana Flávia Machado

Belo Horizonte, MG  
Universidade Federal de Minas Gerais  
Centro de Desenvolvimento e Planejamento Regional  
2007

## Agradecimentos

Em primeiro lugar, gostaria de agradecer à professora Ana Flávia Machado. Sua orientação segue o sentido literal da palavra. Sinto-me muito grato por ter tido a oportunidade de observar seu compromisso, seriedade e empenho. Exemplo de profissionalismo, o qual hei de me recordar ao longo de minha carreira acadêmica.

Em segundo momento, agradeço à professora Sueli Moro, por seus ensinamentos e ajuda com a tarefa de busca de dois dos principais textos abordados nesta dissertação. Também pelas correções finais gentilmente apresentadas.

Devo novamente agradecer à prof<sup>a</sup> Maria da Conceição Sampaio de Sousa (minha orientadora na graduação da UnB), por ter me ensinado os permeandros e alguns dos segredos da intricada, porém bela, microeconomia. O método de mensuração de eficiência veio como consequência de seu apreço pela Economia do Setor Público, o qual, através dela, eu aprendi a compartilhar.

Agradeço aos colegas do Cedeplar pela diversão e cizânias proporcionadas. À família Cedeplar, também, só posso agradecer, por ter fornecido um ambiente de estímulo acadêmico incomparável aos que tive a oportunidade de conhecer pelo país.

Finalmente, agradeço aos meus entes queridos, que me apóiam e dão força, mesmo quando alguns deles não entendem porque a vida acadêmica nos pede um adiamento das vantagens mais imediatas que nos surgem e, às vezes, cheguem a se perguntar ‘pra que é que serve mesmo um economista?’.

# ÍNDICE

1. Introdução.....	10
2. A Economia da Educação.....	17
2.1. Efeitos do desempenho educacional sobre crescimento, desigualdade e bem-estar.....	17
2.2. Desempenho educacional e Eficiência nas escolas.....	22
3. Métodos.....	28
3.1. Método DEA: Data Envelopment Analysis.....	28
3.2. Novas incorporações do método DEA.....	36
3.3. O novo método de dois estágios.....	41
3.4. Detecção de Outliers.....	45
4. Fonte de Dados.....	48
4.1. Dados do SIMAVE.....	48
4.2. Dados do SICA.....	55
4.3. Dados do Censo Escolar.....	59
5. Resultados.....	71
5.1. Eficiência escolar mineira.....	71
<b>5.1.1. Eficiência 4ª série.....</b>	<b>72</b>
<b>5.1.2. Eficiência 8ª série.....</b>	<b>87</b>
<b>5.1.3. Eficiência 3ª série do ensino médio.....</b>	<b>94</b>
5.2. Eficiência por Município.....	102
6. Conclusão.....	107
7. Bibliografia.....	110
Anexo A (Métodos).....	110
Anexo B.....	121
Anexo C.....	123
Anexo D.....	128
Anexo E.....	142
Anexo F.....	152

## ÍNDICE DE TABELAS

Tabela 1.1. Indicadores do Ensino Fundamental, Brasil e Minas Gerais. ....	13
Tabela 1.2. Indicadores do Ensino Médio, Brasil e Minas Gerais. ....	14
Tabela 1.3. Formação dos Professores e Escolas por Esfera Administrativa. ....	15
Tabela 4.1. Resumo dos resultados das provas por anos de Ensino. ....	52
Tabela 4.2. Variáveis do Questionário do SIMAVE. ....	53
Tabela 4.3. Resumo das informações do Questionário do SIMAVE 4 <sup>a</sup> Série. ....	54
Tabela 4.4. Resumo das informações do Questionário do SIMAVE 8 <sup>a</sup> Série. ....	54
Tabela 4.5. Resumo das informações do Questionário do SIMAVE 3 <sup>a</sup> Série. ....	54
Tabela 4.7. Resumo das informações do Censo escolar, variáveis não agrupadas 4 <sup>a</sup> Série. ....	68
Tabela 4.8. Resumo das informações do Censo escolar, variáveis não agrupadas 8 <sup>a</sup> Série. ....	68
Tabela 4.9. Resumo das informações do Censo escolar, variáveis não agrupadas 3 <sup>o</sup> Ano. ....	69
Tabela 4.10. Resumo das variáveis de <i>outputs</i> e <i>inputs</i> do 1 <sup>o</sup> estágio. ....	70
Tabela 4.11. Resumo das variáveis <i>Background</i> e <i>dotação</i> do 2 <sup>o</sup> estágio. ....	70
Tabela 5.1. Média dos Produtos e Insumos por Mesoregião 4 <sup>a</sup> série. ....	72
Tabela 5.2. Ranking por DEA- <i>bootstraped</i> (DEA_c*) para a 4 <sup>a</sup> série: ....	73
Tabela 5.3. Ranking índice Sampaio-Stosic (DEA_ss) para a 4 <sup>a</sup> série: ....	75
Tabela 5.4. Ranking do quociente eficientes/total de escolas 4 <sup>a</sup> série: ....	76
Tabela 5.5. Distribuição das escalas de rendimento por Mesoregião: ....	77
Tabela 5.6. Ranking das mesoregiões para a 4 <sup>a</sup> série: ....	78
Tabela 5.7. Resumo das Variáveis da Função para Escolas Eficientes 4 série: ....	79
Tabela 5.8. Resumo das Variáveis da Função para Escolas NÃO Eficientes 4 série: ..	79
Tabela 5.9. Resultados das Regressões dois estágios 4 <sup>a</sup> série. ....	84
Tabela 5.10. Ranking das Mesoregiões para a 8 <sup>a</sup> série. ....	88
Tabela 5.11. Resultados das Regressões dois estágios 8 <sup>a</sup> série. ....	84
Tabela 5.12. Ranking das Mesoregiões para a 3 <sup>a</sup> série. ....	96
Tabela 5.13. Resultados das Regressões dois estágios 3 <sup>a</sup> série. ....	97
Tabela 5.14. Ranking das cidades com Escolas Eficientes ..... 103	
Tabela 5.15. Distribuição das Mesoregiões nos <i>rankings</i> . .... 104	
Tabela 5.16. Municípios <i>outliers</i> . .... 104	
Tabela 5.17. Resultados da regressão DEA-Municípios. .... 105	
Tabela D.1 Valores dos $d_i$ críticos para os testes de escala. .... 133	
Tabela D.2 Valores dos $l_c$ críticos para teste de detecção de <i>outliers</i> . .... 137	
Tabela D.3 Teste de Aderência entre escolas de Qualidade e a eficiência DEA.* ..... 138	
Tabela E.1. Resumo para Variáveis de Atendimento e Infra-estrutura Eficientes 4 <sup>a</sup> série: .... 143	
Tabela E.2. Resumo para Variáveis de Dotação 4 Série: .... 143	
Tabela E.3. Média dos Produtos e Insumos por Mesoregião 8 <sup>a</sup> série. .... 144	
Tabela E.4. Distribuição das escalas de rendimento por Mesoregião 8 <sup>a</sup> série: .... 144	
Tabela E.5. Resumo Variáveis da Função para Escolas Eficientes 8 série: .... 145	
Tabela E.6. Resumo Variáveis da Função para Escolas NÃO Eficientes 8 série: ..... 145	

<b>Tabela E.7. Resumo para Variáveis de Atendimento e Infraestrutura Eficientes 8ª :</b>	146
<b>Tabela E.8. Resumo Variáveis Dotação 8 Série:</b>	146
<b>Tabela E.9. Média dos Produtos e Insumos por Mesoregião 3ª série.</b>	147
<b>Tabela E.10. Distribuição das escalas de rendimento por Mesoregião 3ª série:</b>	147
<b>Tabela E.11. Resumo Variáveis da Função para Escolas Eficientes 3 série:</b>	148
<b>Tabela E.12. Resumo Variáveis da Função para Escolas NÃO Eficientes 3 série: ...</b>	148
<b>Tabela E.13. Resumo para Variáveis de Atendimento e Infraestrutura Eficientes 3ª série.</b>	149
<b>Tabela E.14. Resumo Variáveis Dotação 3ª série.</b>	149
<b>Tabela E.15. Regressões truncadas para novos indicadores de infraestrutura 4ª Serie.</b>	150
<b>Tabela E.16. Regressões truncadas para novos indicadores de infraestrutura 8ª Serie.</b>	150
<b>Tabela E.17. Regressões truncadas para novos indicadores de infraestrutura 3ª Serie.</b>	151

# ÍNDICE DE GRÁFICOS E FIGURAS

## Gráficos:

Gráfico 3.1. Construção da Eficiência DEA_C.....	31
Gráfico 3.3. Construção da Eficiência DEA_V.....	33
Gráfico 3.4. Densidade da eficiência não corrigida DEA_C.....	39
Gráfico 3.5. Densidade da eficiência corrigida DEA_C.....	42
Gráfico 4.1. Histograma e distribuição para as Notas de Matemática e Português.....	51
Gráfico 4.2. Histogramas e Distribuições dos Custos Dentro por Série.....	57
Gráfico 4.3. Curvas de Custo-Médio por Série.....	58
Gráfico 4.4. Histograma e distribuição de Kernel para <i>qdocentes</i> .....	62
Gráfico 4.5. Histograma para Numero de Salas.....	63
Gráfico 4.6. Histograma para <i>Infra</i> -estrutura.....	64
Gráfico 4.7. Histograma e distribuição Normal para <i>VCM</i> .....	66
Gráfico 4.8. Histograma e distribuição de Kernel para <i>Matrículas por período</i> .....	67
Gráfico 5.1. Ganho nas Notas pelo critério de Eficiência Output 4ª série.....	85
Gráfico 5.2. Ganho na Média de Matrículas pelo critério de Eficiência Output 4ª série.....	86
Gráfico 5.3. Ganho nas Notas pelo critério de Eficiência Output 8ª série.....	92
Gráfico 5.4. Ganho na Média de Matrículas pelo critério de Eficiência Output 8ª série.....	93
Gráfico 5.5. Ganho nas Notas pelo critério de Eficiência Output 3ª série.....	99
Gráfico 5.6. Ganho na Média de Matrículas pelo critério de Eficiência Output 3ª série.....	99
Gráfico 5.7. Comparativo da Evolução das Notas por Série.....	100
Gráfico 5.8. Comparativo da Evolução das Matrículas por Série.....	101
Gráfico A.1. Medição da Eficiência orientação do insumo.....	118
Gráfico A.2. Medição da Eficiência orientação do insumo.....	119
Gráfico C.1. Densidade de Kernel para ‘apvídeo’ (4ª série).....	123
Gráfico C.2. Densidade de Kernel para ‘comp_imp’ (4ª série).....	124
Gráfico C.3. Densidade de Kernel para ‘multimid’ (4ª série).....	125
Gráfico D.1. Histogramas para estatística $d_i$ , todas as séries.....	130
Gráfico D.2. Histogramas da distribuição dos índices de eficiência com os respectivos cortes.....	130
Gráfico D.3. Histogramas da distribuição dos <i>leverages</i> com respectivos cortes.....	130
Gráfico D.4. de Fronteiras Interiores - modelo CFS de exclusão de <i>outliers</i> .....	140
Gráfico F.1. Histograma do $d_i$ para a DEA-municípios.....	152

## Figuras e Mapas:

Figura 2.1. Estimação da Fronteira Condicionada com Erros Normais.....	23
Figura 2.2. Estimação da Fronteira Condicionada com Erros Normais-Truncados.....	24
Figura 3.1. Procedimento para obtenção de rendimentos de escala.....	34
Mapa 5.1. Eficiência Média por Mesoregião de Minas Gerais.....	72

## RESUMO

Esta dissertação 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.

Os métodos utilizados baseiam-se em contribuições recentes dos estudos de eficiência não paramétrica, Simar e Wilson (1998), e semi-paramétrica, Simar e Wilson (2002). Estiveram presentes preocupações para o controle de *outliers* sugeridas pelos trabalhos de Cazals et alii (2002) e Sampaio de Sousa e Stosic (2005). A incorporação dos procedimentos estatísticos de *bootstrap* forneceu mais robustez para os resultados encontrados e permitiu encontrar o viés e intervalos de confiança para os índices de eficiência.

As bases censitárias do Sistema Mineiro de Avaliação da Educação Pública (SIMAVE), que aplica os exames de proficiência de Matemática e Português para 4<sup>a</sup> e 8<sup>a</sup> séries do fundamental e 3<sup>a</sup> 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.

## 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.

All methods discussed here are based on recently papers of non-parametrical analysis, Simar and Wilson (1998), and semi-parametrical analysis, Simar and Wilson (2002). Further analyses concerned about *outlier-detection* are implemented: Cazals et ali (2002) and Sampaio de Sousa e Stosic (2005). Those new methods, based on *DEA-bootstrap*, have improved the robustness of analysis and made became possible an efficiency confidence intervals construction.

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 show inputs complementarity's, in and out school. 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.

# 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 disso, 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 de estudo 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<sup>1</sup> e os resultados, ainda assim, são decepcionantes.

Esta dissertação emprega o método não paramétrico da *Data Envelopment Analysis* (DEA) para medir a eficiência econômica da educação estadual de Minas Gerais. A eficiência econômica é definida como a maneira de obter os maiores ganhos no produto recorrendo ao mínimo possível de insumos. Nos parâmetros educacionais, implica obter o melhor possível em termos do desempenho dos alunos nos exames de proficiência nas matérias básicas dados os recursos que sua escola possui em termos de infra-estrutura, custo por aluno, qualificação de professores, salas de aula e equipamentos.

---

<sup>1</sup> Gasto referente à todos os níveis educacionais para o ano de 2002, segundo dados do “*Education at Glance 2006*” da OCDE: <http://www.oecd.org>. Os gastos do Brasil estão no mesmo patamar que países de desenvolvimento semelhante e até um pouco maior como Austrália, 4,3%, Chile, 3,5%, Irlanda 4,1%, Rússia, 3,7%.

As recentes bases do Sistema Mineiro de Avaliação da Educação Pública (SIMAVE) e do Sistema Informacional Custo Aluno (SICA)<sup>2</sup> ajudaram a compor as informações de produtos e insumos dos modelos. Um questionário sócio-econômico que acompanha a pesquisa do SIMAVE foi muito útil para a composição das variáveis *background*. Além dessas duas bases, empregou-se também o Censo Escolar do MEC/INEP com o fim de complementar as informações de cada escola.

No capítulo a seguir, revisamos a literatura econômica sobre educação e o seu emprego na análise de eficiência por meio da análise envoltória. No capítulo 3, discorremos sobre o método DEA-clássico e o recente método *DEA-bootstrap*, apresentamos, também, o procedimento de dois estágios e as técnicas de identificação de *outliers*. No capítulo 4, descrevemos as três bases de dados principais empregadas, SIMAVE, SICA e Censo Escolar, e, a construção das variáveis do modelo. No capítulo 5, desenvolvemos os dois modelos de análise de eficiência deste estudo. O primeiro se aplica às 3037 escolas estaduais que ministram as 4<sup>a</sup> e 8<sup>a</sup> séries do ensino fundamental e a 3<sup>a</sup> série do ensino médio.<sup>3</sup> O segundo modelo desenvolve uma análise de eficiência por municípios. No capítulo 6, concluímos com os principais resultados da dissertação.

Não só os estudos sobre educação ganharam importância. Os indicadores demonstram que a educação básica brasileira evoluiu consideravelmente ao longo da década de 90 e início dos anos 2000. Em 1991, quando foi realizado o seminário sobre “Qualidade, Eficiência e Equidade na Educação Básica” do IPEA, identificavam-se uma série de problemas para os quais a educação básica no Brasil

---

<sup>2</sup> A fim de evitar uma confusão de conceitos, cabe ressaltar que a palavra “custo” presente na denominação “custo por aluno” do SICA não representa, a rigor, o custo econômico. Não foi realizada uma composição de gastos e custos de oportunidades relativos no emprego dos recursos destinados à educação, portanto, uma melhor denominação seria “gastos por aluno”.

<sup>3</sup> Dessas 3037 escolas, 527 possuem os três ciclos juntos, 1035, possuem apenas o fundamental, 752 ministram o segundo ciclo do fundamental com o ensino médio, 345 são exclusivas para o primeiro ciclo fundamental, 238 exclusivas apenas para o segundo, 105 que possuem apenas o ensino médio e 35 apenas combinam o primeiro ciclo fundamental com o ensino médio.

se encontrava na estaca zero. Atualmente, muitos dos problemas identificados naquele seminário se encontram, se não resolvidos, praticamente sanados.<sup>4</sup>

Em 1995 o INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – passou por uma importante reestruturação que aventou a implementação das novas avaliações dos diversos níveis de ensino: básico, médio e superior e possibilitou também o levantamento de estatísticas mais amplas e confiáveis sobre a educação.

---

<sup>4</sup> Gomes e Sobrinho, orgs. (1992) “Qualidade, Eficiência e Equidade na Educação Básica”. Os problemas identificados no começo da década eram a necessidade de se universalizar o ensino, demanda por um esquema federal de repasse de recursos e sua estruturação, aplicação de merenda escolar, necessidade de distribuição de livros didáticos e algumas outras.

A evolução dos indicadores para a educação fundamental pública pode ser vista na tabela 1.1 abaixo, onde se destacam, de fato, a maior cobertura educacional, os níveis ainda preocupantes de repetência e o abandono e a melhora das condições de infra-estrutura escolar.

**Tabela 1.1. Indicadores do Ensino Fundamental, Brasil e Minas Gerais.**

	1999	2000	2001	2002	2003	2004	2005
<i>Brasil:</i>							
<b>Cobertura do Ensino*</b>	0,99	0,98	0,99	1,00	1,00	1,00	1,00
<b>Merenda Escolar**</b>	0,94	0,94	0,94	0,93	0,95	0,98	0,96
<b>Taxa repetência</b>	0,11	0,11	0,12		0,13	0,14	
<b>Taxa de Abandono</b>	0,12	0,13	0,10		0,09	0,09	
<b>Doc Ensino Superior</b>	0,45	0,47	0,48	0,51	0,56	0,60	0,65
<b>Escolas com Sanitário</b>	0,81	0,83	0,84	0,85	0,89	0,91	0,92
<b>Escolas com Microcomputador</b>	0,14	0,16	0,19	0,23	0,27	0,29	0,32
<i>Minas Gerais:</i>							
<b>Cobertura do Ensino</b>	0,99	0,99	1,00	1,00	1,00	1,00	1,00
<b>Merenda Escolar***</b>					0,95		
<b>Taxa repetência</b>	0,07	0,07	0,07		0,10	0,11	
<b>Taxa de Abandono</b>	0,11	0,08	0,06		0,07	0,07	
<b>Doc Ensino Superior</b>	0,54	0,55	0,57	0,60	0,62	0,63	0,67
<b>Escolas com Sanitário</b>	0,93	0,94	0,93	0,94	0,97	0,99	1,00
<b>Escolas com Microcomputador</b>	0,18	0,22	0,27	0,31	0,36	0,39	0,42

Dados em *Porcentagem*.

\* Taxa de escolarização líquida, a população foi obtida por interpolação geométrica entre 1990 e 2000.

\*\* Para o Ensino básico todo, não só para o Fundamental.

\*\*\* Merenda escolar obtida a partir dos dados do SIMAVE.

FONTE: INEP/MEC <http://www.edudatabrasil.inep.gov.br/> e FNDE para merenda escolar.

Destaca-se, também, a disparidade regional da Educação, evidenciada em parte pelos dados de Minas Gerais. A região Sudeste, da qual o estado faz parte, possui melhores indicadores do que a média nacional, enquanto que o Norte e Nordeste apresentam-se, sempre, com os piores indicadores.

Como bastante ressaltado, o quadro para o ensino médio é mais grave, menor cobertura e maior taxa de abandono. Os indicadores de menor repetência só indicam

que a abstenção foi grande nos níveis anteriores ocorrendo uma perversa “seleção natural”.<sup>5</sup> Embora detenha indicadores de infra-estrutura melhores, a situação se complica para o ensino médio, se considerarmos que a estrutura etária da população brasileira acumula proporção cada vez maior de jovens e adultos, pressionando uma maior demanda sobre o ensino neste nível.

**Tabela 1.2. Indicadores do Ensino Médio, Brasil e Minas Gerais.**

	1999	2000	2001	2002	2003	2004	2005
<i>Brasil:</i>							
<b>Cobertura do Ensino*</b>	0,24	0,24	0,24	0,24	0,25	0,24	0,23
<b>Taxa repetência</b>	0,08	0,08	0,09		0,11	0,11	
<b>Taxa de Abandono</b>	0,18	0,19	0,17		0,17	0,18	
<b>Doc Ensino Superior</b>	0,88	0,88	0,89	0,89	0,90	0,92	0,95
<b>Escolas com Sanitário</b>	0,97	0,97	0,96	0,97	0,99	0,99	1,00
<b>Escolas com Microcomputador</b>	0,67	0,72	0,75	0,80	0,87	0,87	0,89
<i>Minas Gerais:</i>							
<b>Cobertura do Ensino*</b>	0,26	0,28	0,27	0,25	0,26	0,26	0,25
<b>Taxa repetência</b>	0,06	0,07	0,08		0,13	0,12	
<b>Taxa de Abandono</b>	0,26	0,18	0,18		0,16	0,18	
<b>Doc Ensino Superior</b>	0,90	0,91	0,93	0,94	0,94	0,94	0,93
<b>Escolas com Sanitário</b>	0,97	0,97	0,95	0,96	0,99	1,00	1,00
<b>Escolas com Microcomputador</b>	0,75	0,77	0,80	0,84	0,87	0,87	0,87

Dados em *Porcentagem*.

\*Taxa de escolarização líquida, a população foi obtida por interpolação geométrica entre 1990 e 2000.

FONTE: INEP/MEC <http://www.edudatabrasil.inep.gov.br/> e FNDE para merenda escolar.

<sup>5</sup> Ao se analisar os indicadores de repetência e evasão brasileiros há sempre que se considerar o fluxo de progressão pelas séries. Quando o número de crianças matriculadas em uma série de ensino passa de 100% para 50% na série seguinte isso, não necessariamente, implica evasão de metade dos alunos, mas sim que metade dos alunos não tenha passado de ano. As principais conclusões de Fletcher e Ribeiro (1989) apontam que o problema da repetência, e não propriamente o da evasão, é um dos mais sérios. O fato do aluno se atrasar anos repetidos desestimula seu aprendizado e acaba ocasionando uma evasão no futuro por conta de seu insucesso escolar.

A qualificação dos professores da esfera pública por etapa de ensino é evidenciada na tabela 1.3 abaixo. Minas Gerais possui proporção de professores com ensino superior um pouco maior do que para o restante do Brasil, para a 4ª série, o percentual dos dados do ano de 2005 atinge 50% enquanto que no Brasil é de 46%, para as séries entre 5ª e 8ª o diferencial é de apenas 4% a favor de Minas, e, para o ensino médio, o total de professores com superior é dois pontos percentuais menor em Minas Gerais.

O ensino fundamental é basicamente provido pelos municípios, sendo sua importância relativa ainda maior quando se analisam escolas que lidam apenas com o primeiro período (1ª. a 4ª. Série do Ensino Fundamental). No ensino médio, a esfera estadual passa a ser preponderante. As distribuições por esfera administrativa estão na tabela a seguir.

**Tabela 1.3. Formação dos Professores e Escolas por Esfera Administrativa.**

	1º Ciclo Fund	2º Ciclo Fund	Ens Médio
<i>Brasil:</i>			
<b>Prof. Ensino Fundamental Completo</b>	0,01	0,00	0,00
<b>Professores Ensino Médio Completo</b>	0,46	0,18	0,05
<b>Professores Ensino Superior Completo</b>	0,53	0,82	0,95
<b>Escolas Estaduais</b>	0,13	0,38	0,66
<b>Escolas Municipais</b>	0,75	0,43	0,03
<b>Escolas Particulares</b>	0,12	0,19	0,30
<b>Escolas Federais</b>	0,00	0,00	0,01
<i>Minas Gerais:</i>			
<b>Prof. Fundamental Completo</b>	0,01	0,00	0,00
<b>Professores Ensino Médio Completo</b>	0,49	0,14	0,07
<b>Professores Ensino Superior Completo</b>	0,50	0,86	0,93
<b>Escolas Estaduais</b>	0,22	0,53	0,69
<b>Escolas Municipais</b>	0,67	0,30	0,03
<b>Escolas Particulares</b>	0,11	0,17	0,27
<b>Escolas Federais</b>	0,00	0,00	0,01

Dados em *porcentagem*.

\*Taxa de escolarização líquida, a população foi obtida por interpolação geométrica entre 1990 e 2000.

FONTE: INEP/MEC <http://www.edudatabrasil.inep.gov.br/> e FNDE para merenda escolar.

O momento atual no Brasil é propício para as tentativas de resolução das questões de qualidade do ensino fundamental e médio brasileiro. As demandas de primeira necessidade levantadas no início da década se encontram em processo de resolução. Todavia, a medida que os problemas mais básicos são amenizados, revelam-se outras fraquezas de nossa educação, dentre elas se destaca atualmente a baixa qualidade do ensino.

## **2. A Economia da Educação.**

### ***2.1. Efeitos do desempenho educacional sobre crescimento, desigualdade e bem-estar.***

A educação como agenda de pesquisa está consolidada na literatura econômica, por isso, já se identifica uma área de conhecimento denominada **economia da educação**. O número de livros e artigos sobre o assunto é vasto e tais trabalhos analisam sob os mais diversos enfoques o papel da educação na sociedade contemporânea. O interesse na área surge por causa de comprovadas evidências que demonstram que a formação educacional de uma população contribui para o seu desenvolvimento intelectual e material.

A teoria do Capital-Humano, desenvolvida por Becker, nos anos 60, enfatiza a educação como um investimento pessoal que traz retornos futuros dada uma taxa de desconto, sendo possível visar o benefício líquido advindo do aumento de produtividade do trabalho conferido pela educação (MINCER, 1962, BEHRMAN e BIRDSALL, 1983 e GARCIA et al., 2005).<sup>6</sup>

Quanto aos ganhos macroeconômicos de uma população mais bem educada, grande parte da literatura tende afirmar sua significativa importância, tal afirmação é corroborada pelos mais diferentes métodos.<sup>7</sup> Lau, Jamison e Louat (1991) estimam uma função de produção agregada para captar o efeito do Capital-Humano sobre o crescimento de países em desenvolvimento. Utilizando dados de 1960 a 1987, os autores encontram que os efeitos da educação sobre o crescimento são significativos.

---

<sup>6</sup> Uma literatura relacionada que procura identificar como os investimentos em educação se realizam entre gerações são os trabalhos de Becker (1964) e Becker, Murphy e Tamura (1990).

<sup>7</sup> Prichett (1996) escreveu um artigo controverso sobre o papel das escolas e dos ganhos coletivos em Capital Humano, "Where has all the Education Gone?", os argumentos apresentados são que os insumos investidos em educação não têm importância no crescimento econômico, sendo mais importante o lado das capacidades individuais. Krueger e Lindahl (2001) rebatem essa visão, argumentando que é preciso outro procedimento para os dados que tratam de incorporação de Capital-Humano no crescimento.

Nesse estudo, a educação é medida pelos anos de escolaridade média da população, primário e secundário. Para o caso específico da América Latina, concluem que um ano a mais de escolaridade do primário ocasiona o crescimento do PIB à taxa de 2,67% e uma taxa de 3,82% para um ano a mais no nível secundário. Sala-i-Martin (2002) traz uma revisão da nova literatura que incorpora capital humano e variáveis sócio-econômicas nas equações de crescimento. É também estabelecido que o investimento em educação colabora para uma distribuição mais equânime do crescimento econômico, auxiliando na distribuição da renda, e que os gastos públicos com educação devem, portanto, contribuir significativamente para a redução das desigualdades. Para o caso do papel da educação na redução das desigualdades de rendimentos no Brasil ver Barros, Machado e Mendonça (1997).<sup>8</sup>

Os estudos com enfoque nos aspectos macroeconômicos da educação ajudam a estabelecer sua relevância e determinar quais são as conseqüências no produto e no bem-estar de uma população mais bem educada. Porém, a sua generalidade não nos permite identificar muito bem os motivos individuais ou coletivos para se investir em educação. Tendo em vista esse último aspecto, é preciso direcionar o estudo para sua parte microeconômica. Os trabalhos que seguem a linha proposta por Becker ou Mincer enfatizam os ganhos individuais à partir do retorno financeiro de uma maior educação, mas não procuram os fatores que *qualifiquem* a educação. É sabido que a escolarização se diferencia entre os indivíduos, oito anos de educação formal podem não significar o mesmo para duas pessoas que compartilhem características comuns mas que tenham estudado em escolas diferentes, ou então, que tenham estudado na mesma escola mas tenham uma bagagem de educação familiar diferente. Esses são alguns fatores que a agenda microeconômica da educação procura tratar com mais detalhes.

---

<sup>8</sup> Em Barros e Mendonça (1997) se encontram evidências empíricas de que há no Brasil um sub-investimento em educação. Os autores realizam um cálculo comparativo de como esse resultado possivelmente atrasa o crescimento do PIB.

Dentro dessa seara, o ponto de debate mais acirrado se dá no nível de relevância dos insumos. Em 1965, os EUA implementaram o programa *head start*, iniciativa que visava contribuir para o desenvolvimento geral das crianças de mães carentes, mães negras principalmente. Objetivando encontrar evidências corroborando que o desempenho aquém do esperado obtido pelos alunos negros era devido a uma insuficiência de insumos em suas escolas, Coleman et ali (1966) agrupam, pela primeira vez, os dados de educação em uma função de produção para a escola<sup>9</sup>:

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

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’.

Contrariando as expectativas, o relatório Coleman encontrou que o mau desempenho das crianças negras era pouco explicado pela deficiência de insumos em suas escolas. A maior parte se devia ao que aqui chamamos de ‘ $z_i$ ’, as variáveis de *background* dos indivíduos.

Os resultados do Coleman *report* e o uso criativo de uma função de produção escolar fizeram surgir uma série de textos sobre a questão dos insumos educacionais. Os trabalhos de Hanushek (1986) e Hanushek e Luque (2002) empregam exames de proficiência em uma amostra de escolas norte-americanas e, em uma *cross-country* respectivamente, contrastando os resultados das provas com insumos do tipo professor/aluno, salários docente ou tamanho das turmas. Em ambos trabalhos, as conclusões são de que os insumos analisados não surtem efeitos sobre o desempenho educacional, tanto em países desenvolvidos quanto nos em desenvolvimento. “O que importaria, então, para explicar o rendimento dos alunos?”. O próprio Hanushek

---

<sup>9</sup> Para manter a notação empregaremos aqui a mesma taxonomia utilizada na literatura de eficiência, o que não afeta o significado da função de produção.

busca responder a esta questão em um artigo de 1992 onde se volta para as características da família que indicam o tempo dispensado à educação dos filhos como número de filhos, presença e sexo do chefe do domicílio, jornada de trabalho da mãe, divórcio e uma variável não relacionada à família que é a habilidade do professor. Para capturar a habilidade dos professores, Hanushek usa a média dos ganhos de proficiência nos exames.<sup>10</sup> O autor conclui que esta última variável pode ser importante. Além do tamanho pequeno da família, as demais variáveis não foram significativas. A linha de pesquisa de Hanushek sintetiza a vertente da economia da educação que enfoca o papel das características familiares como o cerne do desenvolvimento educacional.

Contrapondo um pouco a abordagem anterior, temos os trabalhos de Birdsall (1985), que estuda o caso de insumos públicos educacionais no Brasil para aumento da oferta educacional, e Card e Krueger (1993), que seguem uma linha onde os insumos educacionais são analisados por indivíduos de uma coorte, o impacto da educação é medido em termos de rendimentos no futuro. Os autores mencionados encontraram significância dos insumos como professor/aluno, horas-aula e salário docência, reforçando o papel da escola.

No Brasil, a literatura sobre educação se expandiu consideravelmente durante a década de 90. No plano de crescimento e redução das desigualdades através da educação, temos os já mencionados Barros et al (1997) e Barros e Mendonça (1997). Barros et al. (2001) enfocam as características familiares e reforçam mais uma vez porque é importante investir em educação no Brasil para reduzir as desigualdades e evitar a armadilha da pobreza propiciada pela educação.

Quanto à análise de indicadores destaca-se o livro organizado por Riani e Rios-Neto (2004) que trata sobre a construção de indicadores demográficos de educação

---

<sup>10</sup> A amostra de Hanushek se aplica à famílias negras de baixa renda observadas entre 1971 e 1975 do *Gary Income Maintenance Experiment*, um grupo que recebia auxílio de renda e outro grupo de controle. O exame de proficiência utilizado é o *Iowa Reading Comprehension and Vocabulary Tests*. Para se inferir a qualidade, o professor deveria ter 3 ou mais alunos das famílias participantes do programa.

freqüentemente utilizados na literatura como: taxa de escolarização bruta e líquida, taxas de rendimento, taxas de transição, condições de oferta e qualificação de docentes. Além disso, aborda os modelos de Profluxo (FLETCHER e RIBEIRO, 1996; e GOLGHER 2004) que controlam as entradas de alunos matriculados e saídas dos evadidos. Ainda nesse livro, conclusões sobre a evolução dos anos de estudo no Brasil podem ser tomadas pelo modelo de Probabilidade de Progressão por Série (PPS), que calcula a probabilidade dos alunos progredirem a cada série e fornece, no final, o número médio de anos de estudo da população.<sup>11</sup> Rios-Neto, César e Riani (2002) mostram que a maior educação dos professores pode substituir os anos de estudo da mãe, o que contribuiria para desarmar a armadilha da pobreza mencionada.

Os estudos com modelos hierárquicos de Albernaz, Ferreira e Franco (2002), Riani (2005) e Machado et al. (2006) reforçam a preponderância do *background* familiar ante aos insumos escolares.

Partindo do ponto em que deve existir uma subestimação do efeito escola quando comparado às variáveis *background*, Felício e Fernandes (2005) realizam a aplicação de dois métodos alternativos. O primeiro capta a desigualdade entre escolas (efeito escola) a partir das notas do SAEB 2001 em Matemática e Língua Portuguesa. A partir de uma decomposição do índice Theil para desigualdade intra-escolas e entre escolas, os autores encontram que o efeito escola para a prova de Matemática varia de um mínimo de 8,7 até o máximo de 34,4%. O segundo método capta o efeito fixo das melhores escolas e constrói um *ranking* obtido pelos coeficientes das *dummies* de efeito escola do primeiro procedimento. O efeito dessas melhores escolas é significativo, mostrando que é possível os alunos evoluírem o equivalente a três anos de estudo nas provas de Matemática e 2,4 para prova de Português.

---

<sup>11</sup> Todos esses indicadores podem ser denotados como indicadores de eficiência, muito embora não seja ainda a eficiência estritamente econômica a ser tratada nessa dissertação.

## **2.2. Desempenho educacional e Eficiência nas escolas.**

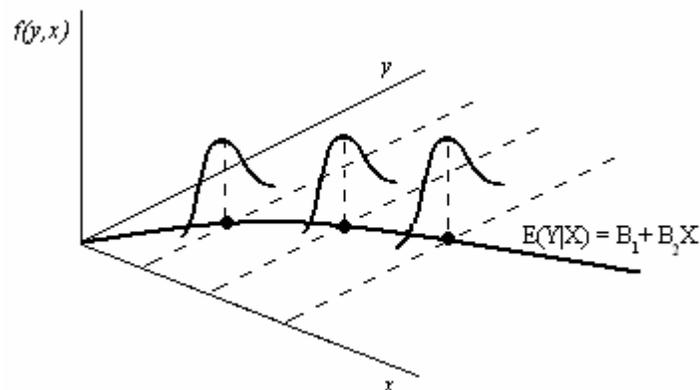
As questões sobre a alocação de recursos são uma das mais difíceis questões práticas com a qual uma sociedade deve lidar em seu dia a dia. É difícil estabelecer a ordem de prioridades, sendo que a mesma passa por constantes mudanças. Por ser tão importante a distribuição de recursos, é preciso que seu uso se dê de forma a promover o máximo de benefício social possível.

A eficiência econômica trata deste aspecto, sua definição é dada como a maneira de conseguir o produto máximo dado determinado volume de recursos. Ou então, estabelecida uma meta para o produto, como conseguiu-la despendendo os menores custos.

Na linguagem da eficiência econômica, a construção teórica importante é a da **função de produção** ou **fronteira de eficiência**. Usa-se, portanto, a função de produção educacional abordada na primeira sub-seção deste capítulo.

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 2.1.

**Figura 2.1. Estimação da Fronteira Condicionada com Erros Normais.**



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 se 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  $\theta_i$ , 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  $\theta_i$  inferior à unidade indica que a observação produz aquém do que é possível.

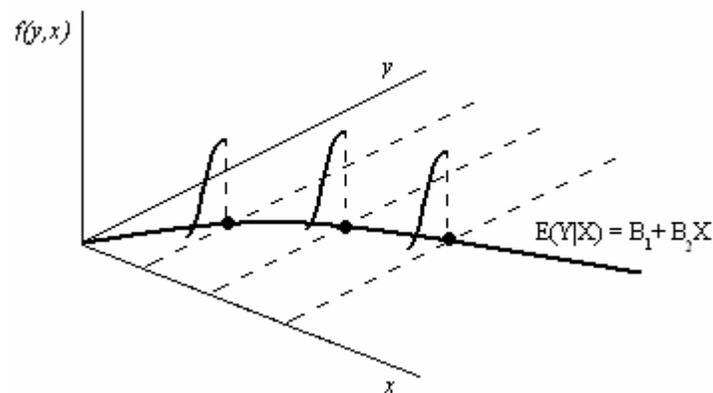
$$y_i = \theta_i f(x_i, z_i, d_i) \quad (2)$$

A partir desse momento, o problema passa a ser como obter a estimação da função sem considerar os pontos estimados como eficientes na média. Uma maneira paramétrica de se conseguir isso é estabelecer a distribuição dos erros como uma normal truncada de erros não-positivos, método conhecido como *fronteiras estocásticas*.<sup>12</sup> O usual é decompor o erro em duas partes: a parte dos erros normais ( $v_i$ ) e a parte dos erros não negativos ( $u_i$ ). Uma equação para estimação da fronteira pode ser especificada da seguinte forma:

$$\ln(y_i) = x_i \beta + v_i - u_i \quad (3)$$

Onde  $\ln(y_i)$  é o produto logaritmicado,  $x_i$  as variáveis para os inputs,  $v_i$  os erros aleatórios de acordo com uma distribuição normal e  $u_i$  erros não positivos especificados a partir de uma normal truncada ou de uma exponencial. A evolução dos estudos em fronteiras estocásticas acompanha os trabalhos de Aigner e Chu (1968), Aigner, Lovell e Schmidt (1977) e Greene (1990) e a intuição pode ser acompanhada pela figura abaixo:

**Figura 2.2. Estimação da Fronteira Condicionada com Erros Normais-Truncados.**



<sup>12</sup> Breve explicação dos modelos de fronteiras estocásticas é fornecida no ANEXO A sobre desenvolvimento dos métodos.

Outra maneira de obtenção da fronteira é a estimação não-paramétrica, que não envolve a utilização de 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).

Charnes, Cooper e Rhodes argumentam em seu *paper* de 1978 que a DEA pode ser um procedimento bastante útil para a análise da eficiência no setor público. 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.

O artigo de Charnes et al (1981), além de pioneiro, traz importantes contribuições teóricas para a comparação de programas governamentais utilizando a DEA. Os autores lançam mão do método para analisar a eficiência das escolas presentes no programa *follow through* norte-americano. O *follow through* é um programa que se seguiu ao projeto *head start* do já reportado relatório Coleman. A identificação de escolas carentes no relatório fez surgir um programa de acompanhamento com vistas a acelerar o desenvolvimento das escolas atrasadas. Os autores concluem que o *follow through* não incrementa a eficiência e que talvez o ideal fosse uma combinação entre alguns fatores presentes nas escolas alvo e outros fatores das escolas não participantes.<sup>13</sup>

As contribuições na área de eficiência educacional não paramétrica surgem novamente no trabalho de Fare Grosskopf e Weber (1989). Os autores avaliaram as escolas do Estado norte-americano do Missouri, ano letivo 1985-1986. Alunos da 8ª. série realizaram o *Basic Essential Skills Test* (BEST), que discrimina se os alunos passaram ou não (acertaram mais de 75%) nos testes de Leitura, Matemática e

---

<sup>13</sup> Importante ressalva dos autores é a de que o *follow through* segue o formato distrital da maioria dos programas nos EUA, cada distrito implementa a diretriz nacional com suas modificações. As instituições privadas que financiavam o *follow through* diferenciavam de estado para estado. O programa perdeu sua força e o apoio privado ao longo dos anos, sofrendo um decréscimo no número de matrículas atendido o que pode explicar sua eficiência não comprovada.

Economia e Governo. Para os insumos, foram utilizados o número de alunos que realizaram o teste, as despesas líquidas por distrito, a valorização da área do distrito escolar (semelhante à ‘dotação’) e o número de professores na oitava série. O aspecto mais importante do trabalho de Fare, Grosskopf e Weber não é o uso de *outputs* de proficiência, medição de uso regular nos EUA, e sim a incorporação, pela primeira vez na função educacional, de retornos variáveis de escala. Os artigos anteriores trabalhavam com a hipótese restrita de retornos constantes. Além disso, o trabalho utiliza o método de *jackknifing* para detecção de *outliers* influentes, importantes para a pequena amostra de distritos escolares do Missouri, já que se sabe que a DEA para retornos variáveis possui convergência ainda mais lenta do que com retornos constantes.<sup>14</sup>

O trabalho de McCarty e Yaisawargn (1993) utiliza dois modelos de medição da eficiência para escolas no distrito de Nova Jérsei, EUA. No primeiro modelo é desenvolvido o procedimento de *dois-estágios*, onde são identificadas variáveis sob controle das escolas: proficiência, gastos, relação professor/aluno e variáveis que não estão sob a influência da direção como as variáveis sócio-econômicas que são chamadas de *uncontrolled*. As primeiras variáveis fazem parte da medição de eficiência obtida pela DEA, as demais são regredidas em um modelo tobit. O segundo modelo adotado por estes autores incorpora todas as variáveis dentro da função estimada pela DEA.<sup>15</sup>

Os resultados de McCarty e Yaisawargn corroboraram as evidências que já haviam sido demonstradas pelo departamento de educação de Nova Jérsei, quais sejam, de que existiam escolas distritais ineficientes, não se tratando ser a melhoria da qualidade da educação naquele distrito uma questão de aumentar o volume de recursos e sim de utilizá-los de maneira mais proveitosa.

---

<sup>14</sup> Simar e Wilson (2002).

<sup>15</sup> Daremos mais detalhes sobre o procedimento de dois estágios e a incorporação das variáveis  $z_i$  assim como os novos avanços do método presentes em Simar e Wilson (2003), no próximo capítulo.

Recentemente, dois trabalhos de aplicação da DEA à eficiência da educação se destacam: Wilson (2005a) 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*”<sup>16</sup> – e variáveis não-discrecionárias como status sócio-econômico e escolaridade dos pais. Wilson (2005) incorpora todas as variáveis em uma só estimação da DEA, enquanto que Afonso e Aubyn (2005) desenvolvem o método de dois estágios da maneira antiga e com as novas incorporações. Por serem as primeiras aplicações dos métodos sugeridos por Simar e Wilson (2000 e 2003), ambos os trabalhos não vão muito longe em suas conclusões. Wilson enfoca a eficiência nos países do leste europeu e aponta algumas direções importantes. Afonso e Aubyn concluem pela importância das *environment variables*, como PNB e escolaridade dos pais, para explicar a eficiência educacional dos países.

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, os de Sampaio de Sousa e Ramos (1999, 2000) onde o enfoque sobre a educação é apenas indireto, pois o trabalho se concentra na eficiência dos gastos públicos em geral, e o trabalho de Faria e Januzzi (2006) sobre eficiência de gastos na área de educação e saúde do municípios do Rio de Janeiro. Para o caso da educação básica e utilizando novos métodos, esta dissertação é pioneira.

---

<sup>16</sup> Exame aplicado entre os 28 países membros da OCDE (*Organization for Economic Cooperation and Development*) 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.

## 3. Métodos

### 3.1. 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 agora nos anos 2000 com as incorporações de Gibels et al. (1999) Kneip et al. (2003) e Simar e Wilson (1998, 2003 e 2005). No entanto, seu referencial teórico inicial surge em Debreu (1951).

A partir desse marco teórico, Farrel (1957) praticamente constrói o novo método não paramétrico da DEA, embora o termo *Data Envelopment Analysis* tenha sido cunhado por Charnes, Cooper e Rhodes (1978). Estes últimos autores restabeleceram o uso empírico do método, aplicando-o a análise de eficiência de recursos públicos.

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<sup>17</sup>; 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**  $P$  dado por:

---

<sup>17</sup> As incorporações a serem mencionadas neste capítulo solucionam em grande monta estes problemas, o uso do *bootstrap* auxilia principalmente o tratamento e inferências obtidas a partir de pequenas amostras. Além disso, a amostra utilizada não pode ser considerada pequena, portanto, a aplicação das novas correções sobre uma amostra maior tornará os resultados ainda mais confiáveis.

$$P = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \text{ pode produzir } \mathbf{y}\}; P \subset \mathfrak{R}^{S+M} \quad (4)$$

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) do capítulo 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 pela DEA ou pelo método de fronteiras estocásticas.<sup>18</sup> 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 seja, a fronteira é delimitada pela melhor prática no esquema de produção em voga.

---

<sup>18</sup> 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 como a distribuição uniforme, normal, exponencial e os demais processos.

A construção de tal fronteira exige um método de programação linear que trabalha com o seguinte processo:

$$\theta_i = \theta_i(\mathbf{x}_i, \mathbf{y}_i | P) \equiv \max \{ \theta_i | (\mathbf{x}_i, \theta \mathbf{y}_i) \in P, \theta_i > 0 \} \quad (5)$$

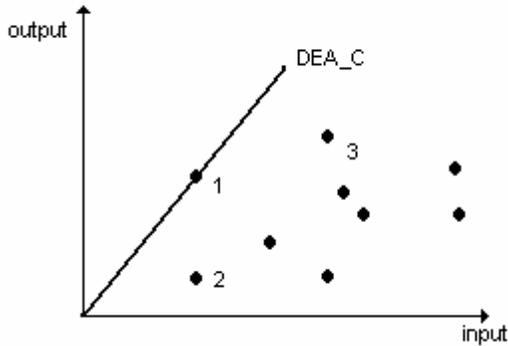
$$sa.P(\mathbf{x}_i, \mathbf{y}_i | C, D) = \left\{ (\mathbf{x}_i, \mathbf{y}_i) \left| \theta_i \mathbf{y}_i \leq \sum_{i=1}^n \lambda_i \mathbf{y}_i, \mathbf{x}_j \geq \sum_{i=1}^n \lambda_i \mathbf{x}_i, \lambda \in \mathfrak{R}^n, \mathbf{y} \in \mathfrak{R}^s, \mathbf{x}_i \in \mathfrak{R}^m \right. \right\} \quad (6)$$

A equação (5) revela um procedimento de maximização onde  $\theta_i$  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 (6) reúne as restrições para a maximização presente em (5). Há alguns novos elementos na equação (6), 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  $\lambda_i$  é um vetor de intensidade,  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N) \in \mathfrak{R}_+^N$ , os  $\lambda$ '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 a fronteira de eficiência representada na figura abaixo. A fronteira indicada se assenta em rendimentos constantes de escala, livre disponibilidade de insumos e na convexidade de fatores, e está representada no  $\mathfrak{R}^2$  (um *output*, um *input*) por simplificação.

**Gráfico 3.1. Construção da Eficiência DEA\_C.**



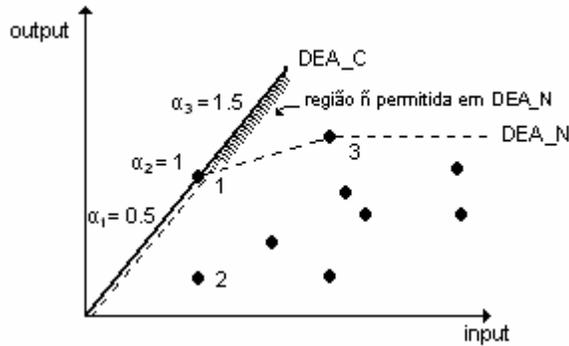
Os pontos representam as observações  $(x_i, y_i)$  e no caso da DEA-C, a fronteira será sempre uma reta, um plano ou hiperplano. A hipótese de retornos constantes pode ser alterada para a possibilidade de retornos não crescentes, 'N':

$$\theta_i = \theta_i(\mathbf{x}_i, \mathbf{y}_i | P) \equiv \max \{ \theta_i | (\mathbf{x}_i, \theta \mathbf{y}_i) \in P, \theta > 0 \} \quad (7)$$

$$sa.P(\mathbf{x}_i, \mathbf{y}_i | N, D) = \left\{ (\mathbf{x}_i, \mathbf{y}_i) \mid \theta \mathbf{y}_i \leq \sum \lambda_j \mathbf{y}_j, \mathbf{x}_i \geq \sum \lambda_j \mathbf{x}_j, \sum \lambda_j \leq 1, \lambda \in \mathfrak{R}^n, \mathbf{y} \in \mathfrak{R}^s, \mathbf{x}_i \in \mathfrak{R}^m \right\}$$

A nova restrição  $\sum \lambda_i \leq 1$  não permite expansões radiais por um escalar 'α' maior do que 1. O gráfico abaixo exemplifica esta construção, comparando-a com a fronteira DEA-C.

**Gráfico 3.2. Construção da Eficiência DEA\_N.**



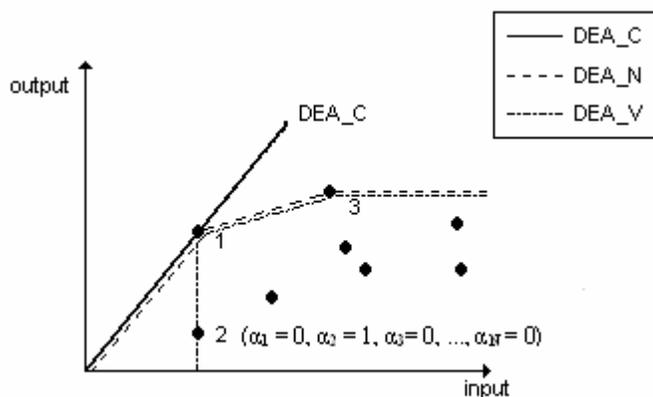
A antiga fronteira DEA-C possuía apenas um ponto eficiente onde  $\alpha_1 = 1$  (ponto 1 na figura), caminhando para a esquerda e para baixo nesta fronteira temos escalares  $\alpha_1 < 1$ , que são permitidos pelo DEA-N. Subindo à direita, a partir do ponto eficiente 1 temos  $\alpha_1 > 1$ , restrição não válida para o DEA-N, o que nos dá a área não permitida assinalada. A terceira possibilidade são os rendimentos variáveis de escala,  $V$ , obtidos por meio da equação (8):

$$\theta_i = \theta_i(\mathbf{x}_i, \mathbf{y}_i | P) \equiv \max \{ \theta_i | (\mathbf{x}_i, \theta \mathbf{y}_i) \in P, \theta > 0 \} \quad (8)$$

$$sa.P(\mathbf{x}_i, \mathbf{y}_i | V, D) = \{ (\mathbf{x}_i, \mathbf{y}_i) | \theta_i \mathbf{y}_i \leq \sum \lambda_j \mathbf{y}_j, \mathbf{x}_i \geq \sum \lambda_j \mathbf{x}_j, \sum \lambda_j = 1, \lambda \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^s, \mathbf{x}_i \in \mathbb{R}^m \}$$

A partir da nova restrição,  $\sum \lambda_i = 1$ , a soma dos escalares está restrita à unidade, indicando que a fronteira está construída aos pares de pontos eficientes. Os pontos se reportam sempre a um segmento de fronteira ( $\lambda_i + \lambda_j = 1$ , onde  $i \neq j$ ). Temos assim o gráfico 3.3 da DEA-V:

**Gráfico 3.3. Construção da Eficiência DEA\_V.**



Observamos que no ponto 2 (agora eficiente) os escalares são  $(\alpha_1 = 0, \alpha_2 = 1, \alpha_3 = 0, \dots, \alpha_N = 0)$ . O que significa que a combinação para este ponto ser eficiente se remete a ele próprio. O mesmo vale para os outros pontos eficientes da fronteira.

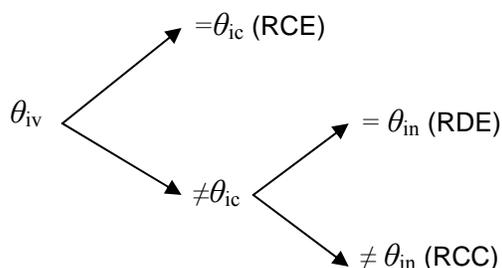
Obtemos com as equações de restrição (6) (7) e (8) a DEA-C, DEA-N e DEA-V. Observamos que o envelope da DEA-V é o mais maleável, envolve os dados a uma distância menor do que as outras duas medidas, isso se refletirá no índice-Farrel de eficiência que, por construção, possui média mais alta. As três formas de medição são úteis para a obtenção dos rendimentos de escala das observações eficientes.<sup>19</sup>

A fronteira DEA-V possui ao menos um ponto em comum com a fronteira DEA-C, como o ponto 1 exemplificado na figura 3.3. Isso indica que quando  $\theta_{iDEA-V} = \theta_{iDEA-C}$ , tem-se rendimentos constantes de escala (RCE). Se não são iguais, cabe verificar duas outras possibilidades: se  $\theta_{iDEA-V} = \theta_{iDEA-N}$ , variável e não crescente, o que indica rendimentos decrescentes de escala (RDE); ou se  $\theta_{iDEA-V} \neq \theta_{iDEA-C} \neq \theta_{iDEA-N}$ , caso onde resta apenas a possibilidade de retornos crescentes (RCC).

<sup>19</sup> 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.

A figura abaixo resume a lógica deste pequeno procedimento que ajuda a encontrar os tipos de rendimentos observados:

**Figura 3.1. Procedimento para obtenção de rendimentos de escala.**



A construção dos índices é feita radialmente a partir da origem e em relação a fronteira. Como observamos, duas formas de construção são possíveis: a eficiência *input-oriented* e a *output-oriented*. Sob retornos constantes de escala, as duas modalidades fornecem os mesmos resultados nos índices de eficiência, e resultados diferentes sob retornos não-crescentes ou variáveis. No entanto, qualquer que seja o retorno de escala empregado, o conjunto de observações identificado como eficientes será o mesmo, e o *ranking* de eficiência permanece também inalterado para qualquer uma das suposições de orientação.<sup>20</sup>

Neste trabalho, utilizaremos o índice de eficiência medido pelo critério de Sheppard (1970) orientado pelo produto (*output-oriented*) e denotado pelo  $\delta_i$ . 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.<sup>21</sup>

<sup>20</sup> Coelli et al. (1998) ressaltam que a escolha entre *input-oriented* e *output-oriented* não é crucial, pois apenas mudam o valor medido em cada um dos índices. A primeira forma está relacionada à minimização de custos enquanto que a segunda se relaciona à maximização de lucros.

<sup>21</sup> A adoção do **Sheppard-eficiência** em *output-oriented* baseia-se no fato de que nessa forma de medição podemos comparar o índice de eficiência à 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**. Detalhes sobre a construção dos índices são fornecidos no ANEXO A.

Ao longo das explicações anteriores evocamos algumas propriedades microeconômicas requeridas pela DEA. Enunciamos, novamente, as três mais importantes: 1<sup>a</sup>.) é o conjunto de observações  $L_N = \{(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i, \mathbf{d}_i)\}$ , que são realizações de processos *iid* com funções densidade probabilidade que nos fornecem  $P$ ; a 2<sup>a</sup>.) é a livre disponibilidade de insumos. A rigor: se  $(x, y) \in P$  e  $\tilde{x} \geq x$  então  $(\tilde{x}, y) \in P$ , o mesmo para  $\tilde{y} \leq y$  que implica  $(x, \tilde{y}) \in P$ ; e 3<sup>a</sup>.)  $P$  é convexo e fechado para todo  $(x, y) \in \mathfrak{R}_+^{M+S}$ , os conjuntos de *inputs* e *outputs* também são convexos e fechados.<sup>22</sup> Ao longo destes últimos anos, uma série de avanços teóricos foram estabelecidos para as propriedades estatísticas dos estimadores não-paramétricos. A sub-seção a seguir apresenta estes novos avanços ao método.

---

<sup>22</sup> Sheppard (1970) fornece propriedades econômicas aplicadas aos estudos de produtividade e eficiência aqui aplicadas e mais algumas que podem ser empregadas de acordo com a ocasião.

### 3.2 Novas incorporações do método DEA.

Banker (1993) apresentou um *paper* onde demonstra, para pequenas amostras, a existência de um viés na fronteira estimada, tal viés se aproxima de zero a medida que a base observada aumenta de tamanho. A partir destas evidências, Banker sugere dois testes de hipóteses baseados na distribuição assintótica dos estimadores. O primeiro teste baseia-se em uma distribuição exponencial para a eficiência, de maneira que, para verificar se dois grupos de índices de eficiência ( $m_1$  e  $m_2$ ) seguem padrões de fronteira diferentes, pode-se utilizar a seguinte fórmula:

$$\left[ \sum_{j \in M_1} \hat{\delta}_j / m_1 \right] / \left[ \sum_{j \in M_2} \hat{\delta}_j / m_2 \right], \text{ que possui distribuição } F \text{ com } (2m_1, 2m_2) \text{ graus de}$$

liberdade. O segundo teste é para o caso de uma distribuição normal truncada, onde temos então  $\left[ \sum_{j \in M_1} (\hat{\delta}_j)^2 / m_1 \right] / \left[ \sum_{j \in M_2} (\hat{\delta}_j)^2 / m_2 \right]$  com  $F(m_1, m_2)$ .<sup>23</sup> Os testes propostos por

Banker foram importantes por estabelecerem os primeiros caminhos a fim de desvendar as propriedades da DEA em pequenas e grandes amostras.

Desenvolvendo o trabalho de Banker, Gibels et al.(1997) estabelecem a distribuição assintótica para a DEA-eficiência no caso de um *input* e um *output*. Kneip et al (2003) trabalham as propriedades assintóticas dos estimadores sob o uso de testes de *bootstrap*, o que consiste em uma prova não-analítica para o caso de múltiplos *inputs* e *outputs*, já que não há ainda prova formal para este último caso.

Para os avanços aplicáveis aos trabalhos empíricos temos três importantes *papers* de Simar e Wilson (1998, 2002 e 2003). O primeiro artigo propõe o uso de dois algoritmos de cálculo da eficiência por *bootstrap*, aqui empregados para a eficiência das escolas estaduais mineras. O ponto inicial é que o conjunto  $P$ , mencionado na

---

<sup>23</sup> Nas fórmulas representadas, Banker utiliza asterisco (\*) para indicar a otimalidade do índice de eficiência, ao contrário do uso para denotar *bootstrap* aqui empregado. O autor usa circunflexo no mesmo sentido que também empregamos: para observações presentes na base e as estimações a partir dela, no entanto, depois da fase de programação linear, Banker usa o asterisco.

seção anterior, é desconhecido e as propriedades assintóticas de seus estimadores são impossíveis de serem obtidas analiticamente. O uso do *bootstrap* permite fazermos inferências com a DEA a partir de propriedades microeconômicas e estatísticas conhecidas e obter uma prática confiável de estimação dos parâmetros, embora não analítica.<sup>24</sup> O artigo de 2002 consolida os testes para a identificação dos rendimentos de escala, importantes para identificar como a unidade de observação pode aperfeiçoar sua produção. Neste artigo, os autores retornam aos testes propostos por Banker e analisam o desempenho de outros testes melhores para captar retornos de escala, baseados na técnica do artigo de 1998.<sup>25</sup> O artigo de 2003 trata do procedimento de dois-estágios a ser abordado na seção 3.3.

Dadas as propriedades e o comportamento que conhecemos da função de produção  $P$ , é possível obter uma estimação de fronteira pelo método DEA. A fronteira estimada é representada por  $\hat{P}$ . O  $\hat{P}$  fornece correção para a base de *inputs* e *outputs* empregada:  $L_{N(x,y)}^* = \{(x_i^*, y_i^*)\}$ . Temos assim, o novo índice de eficiência  $\delta_i^*$  e no processo de obtenção de  $\hat{P}$  é empregado o *bootstrap*, gerando sub-amostras a partir da base inicial,  $L_b^*$ , onde  $b = 1, \dots, B$ .

---

<sup>24</sup> Simar e Wilson (1998, pp. 55) chamam atenção para os cuidados na realização da nova estimação proposta.

<sup>25</sup> Detalhes ANEXO C.

Como ressaltam Simar e Wilson (1998), os dados observados  $\hat{P}$  e sua reamostragem por *bootstrap* devem compor uma representação fiel do verdadeiro  $P$ . O viés que é obtido com o *bootstrap* deve se aproximar do verdadeiro viés:

$$\left(\hat{\delta}_i^* - \hat{\delta}_i\right) \Big| \hat{P} \sim \left(\hat{\delta}_i - \delta_i\right) \Big| P \quad (9)$$

O viés obtido por *bootstrap* é portanto:

$$viés = E_{\tilde{P}}\left(\hat{\delta}_i^*\right) - \hat{\delta}_i = \bar{\delta}_i^* - \hat{\delta}_i \quad (10)$$

Onde  $\hat{\delta}_i^*$  é o Sheppard-eficiência *bootstrap* estimado,  $\hat{\delta}_i$  é a eficiência estimada observada e  $\delta_i$  é a eficiência verdadeira.

Por uma propriedade ainda não muito bem conhecida dos estimadores de eficiência não-paramétricos, a distribuição possui uma massa de probabilidade maior do que zero em torno de 1 (ponto de eficiência), o que torna o viés em (10) negativo.<sup>26</sup> Um exemplo de densidade da distribuição com estimadores não corrigidos está na figura 3.6. E o enunciado da propriedade que se assume para esta distribuição logo em seguida.

---

<sup>26</sup> Simar (2003) aponta que tal ocorrência pode advir das propriedades de amostra finita.

Gráfico 3.4. Densidade da eficiência não corrigida DEA\_C.

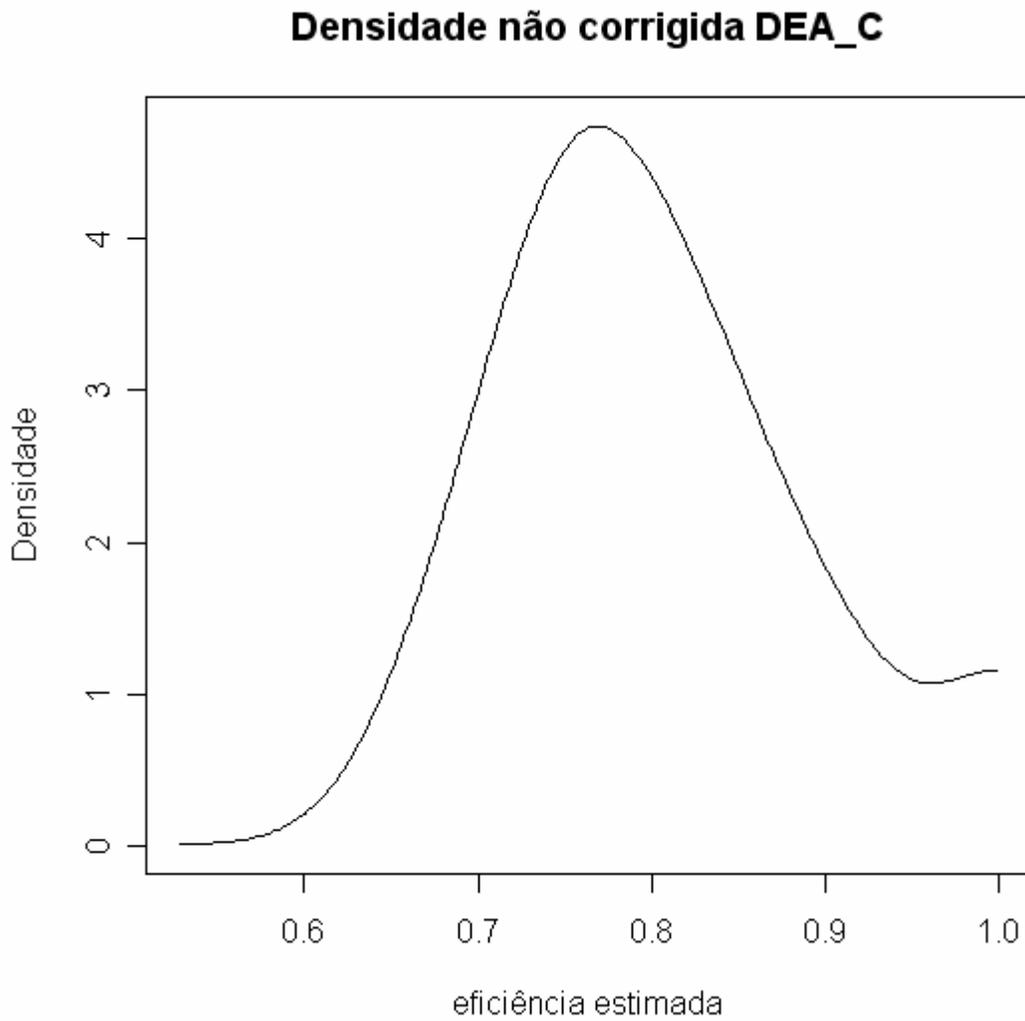


Figura 3.6 - Densidade DEA\_C não corrigida. O gráfico representa o exemplo dos resultados da 8a. série em Minas Gerais

⇒ A densidade  $f(x, y)$  é contínua, exceto sobre a fronteira. Temos  $f(x, y) = 0 \quad \forall (x, y) \in P$  e  $f(x, y) > 0 \quad \forall (x, y) \in P^d$ , o que equivale a dizer que existe uma “corcova” em torno da eficiência acima demonstrada.

O estimador corrigido e o seu desvio-padrão *bootstrap* são os seguintes:

$$\tilde{\delta}_i = \hat{\delta}_i - \text{viés} = 2\hat{\delta}_{i,b}^* - \bar{\delta}_i^* \quad (11)$$

$$s\hat{e} = \left\{ \frac{1}{B-1} \sum_{b=1}^B (\hat{\delta}_{i,b}^* - \bar{\delta}_i^*)^2 \right\}^{1/2} \quad (12)$$

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.

Os avanços recentes resultaram também em mudanças nas práticas estabelecidas dos métodos de dois-estágios e detecção de *outliers* a serem tratados nas duas próximas sub-seções.

### 3.3. O novo método de dois estágios.

A idéia do método de dois estágios é obter duas etapas de estimação: a primeira é a eficiência não-paramétrica da DEA; a segunda é a estimação por MQO, ou outra técnica, dos parâmetros de variáveis que podem influir nos índices de eficiência obtidos no primeiro estágio. Por conta da combinação das duas técnicas, costuma-se denominar por semi-paramétrico o modelo em dois estágios.

A regressão mais simples dos modelos de dois-estágios envolve o índice de eficiência  $\delta_i$  e as variáveis do ambiente, destacadas aqui como  $z_i$ :

$$\delta_i = \beta z_i + \varepsilon_i \quad (13)$$

O  $\beta_i$  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. Entretanto, 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  $\varepsilon_i$ 's são erros da regressão, onde  $E(\varepsilon_i) = 0$  e  $E(\varepsilon_i^2) = \sigma^2$ .

Pela distribuição dos estimadores mostrada na figura (3.6) acreditava-se que o método de estimação correto para a equação (16) era o **tobit**,<sup>27</sup> com  $\delta_i$  censurado à direita em  $\delta_i = 1$ . No entanto, não existia nenhuma maior fundamentação sobre  $\delta$  para explicar porque o método tobit seria o correto. De fato, Simar e Wilson (2003) demonstram por simulações de **monte-carlo** que os intervalos do tobit se situam longe dos valores dos  $\beta$ 's “verdadeiros” estimados por *bootstrap*.

---

<sup>27</sup> O modelo **tobit** é empregado nos casos onde uma das variáveis é censurada ou ausente, ocorrendo um valor maior do que o esperado, viesando a regressão MQO comum.

Outra limitação do modelo tobit segue para o caso onde há correlação entre as variáveis  $z_i$  e os *inputs* e *outputs* do modelo DEA. A estimação MQO se torna viesada, enquanto que a realização do *bootstrapping* para os índices  $\delta_i$  praticamente elimina essa possibilidade. A distribuição corrigida pelo *bootstrap* está representada abaixo. O “calombo” em torno de 1 não está presente e observa-se também uma distribuição mais simétrica com média ligeiramente menor (devido a correção do viés) do que a distribuição não corrigida.

**Gráfico 3.5. Densidade da eficiência corrigida DEA\_C.**

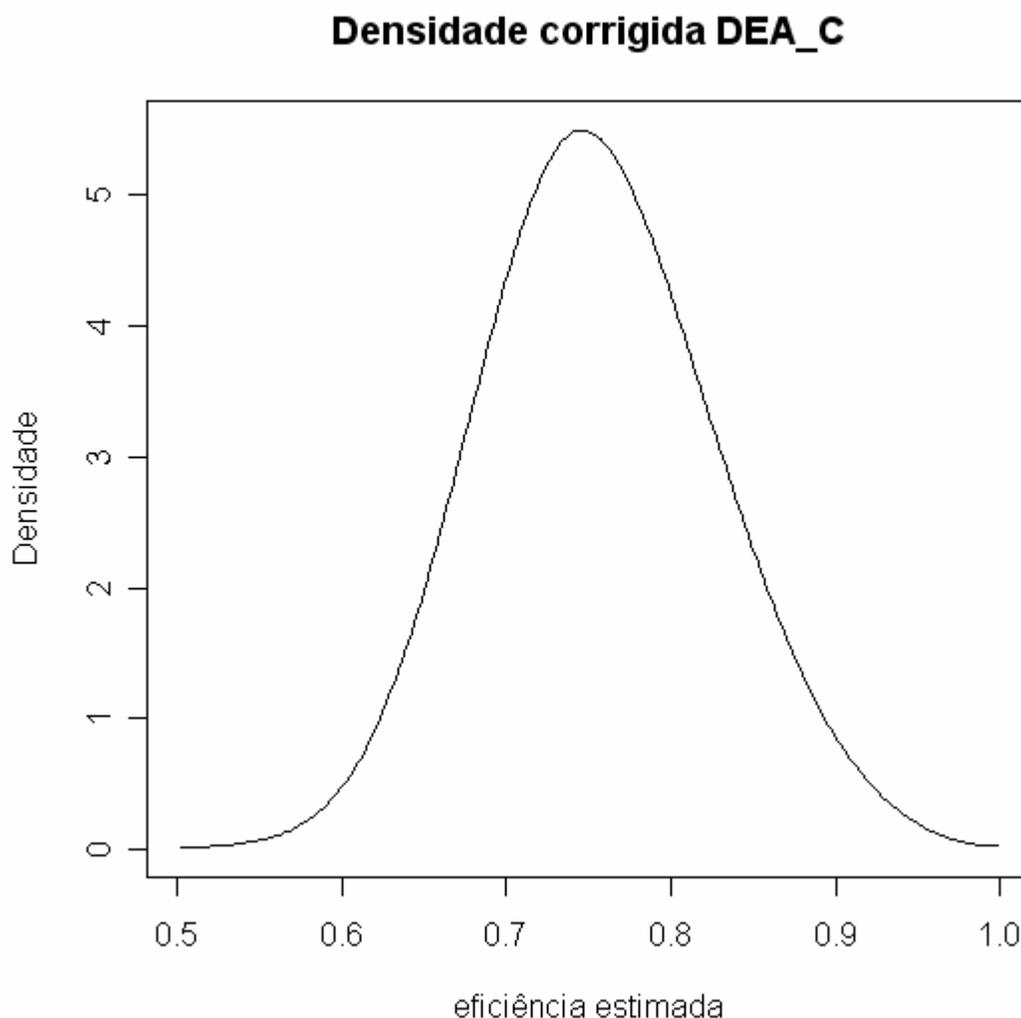


Figura 3.7 - Densidade DEA\_C corrigida para a eficiência da 8a. série em toda Minas Gerais

Não há erros aleatórios na obtenção da DEA, sendo assim, não é difícil perceber que os índices de eficiência estão todos de alguma forma “correlacionados” entre si, uma observação ou é eficiente, ou se reporta a, no mínimo, duas outras observações que formam a fronteira. A correlação entre os  $\delta_i$ 's nos impede de regressar uma equação simples da forma exposta em (13), os  $\varepsilon_i$ 's denotariam a autocorrelação dos índices e a estimação seria menos estatisticamente eficiente. É necessária uma correção, a pesquisa identificava esse problema, o tobit não se tratava de uma correção perfeita, os avanços foram se dando até chegar na idéia de *bootstrap* da fronteira não paramétrica.

Em síntese, a idéia da aplicação dos modelos *bootstrap* é obter maior robustez nos métodos não-paramétricos. Observar o comportamento da fronteira não apenas com a base de dados inicial, mas capturar, por meio desta, sua distribuição provável e conseguir, dessa maneira, 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  $\delta_i$ 's, exclui-se também a parte aleatória do erro que pode ocasionar uma ineficiência.<sup>28</sup> Cada observação permanece, portanto, apenas com a parte que confere ineficiência devido ao mau uso dos fatores. A equação (16) é reescrita da seguinte maneira:

$$\delta_i^* = \beta \mathbf{z}_i + \varepsilon_i \quad (14)$$

O índice de eficiência  $\delta^*$  é o **Sheppard-Eficiência** obtido por *bootstrap*, suas propriedades assintóticas foram tratadas na seção anterior. Ao utilizá-lo, eliminamos a correlação das variáveis  $\mathbf{z}_i$  com  $\delta^*$ . No capítulo 5, mostraremos os passos para a obtenção do índice *bootstrap*, assim como os resultados a partir dos dados das escolas.

---

<sup>28</sup> Dada a distribuição do DGP descrita, os estimadores de eficiência-*bootstrap* passam a ter um componente aleatório e não tão determinístico para sua distribuição.

Além das considerações acima, por ser determinístico, o método DEA está sujeito a alterações bruscas da eficiência devido à observações *outliers*, as chamadas **super-eficientes**. A ocorrência destes casos desloca toda a fronteira e prejudica a análise das demais observações. O desenvolvimento teórico recente preocupou-se também com esta questão e os métodos para detecção de *outliers* são tratados na sub-seção seguinte.

### 3.4. Detecção de Outliers

Não existe definição sedimentada na estatística para descrever o que seria um *outlier*, entretanto, há um consenso de que um *outlier* é uma informação discrepante, que possui um DGP diferente do qual se apresenta para o restante dos dados.<sup>29</sup>

Os “*outliers*” devem ser uma preocupação em qualquer modelo estatístico por viesarem a inferência amostral, em especial nos modelos não-paramétricos onde os dados possuem papel de destaque ao ditar o padrão estatístico que a estimação segue.

Um importante trabalho inicial para a detecção de *outliers* é apresentado por Wilson (1995). O método proposto se baseia no uso de funções de influência. No entanto, o teste de Wilson é computacionalmente intensivo, o que prejudica seu uso em grandes amostras, como a que empregamos neste trabalho.

Seguindo o propósito de Wilson (1995) de detectar e excluir as observações mais influentes, mas com o intuito de aplicar uma correção de *outliers* útil para grandes amostras, Sampaio de Sousa e Stosic (2005) propõem o método “jackstrap” (combinação de “*jackknifing*” com “*bootstrap*”) e o aplicam para a mensuração de eficiência em gastos públicos municipais do Brasil.

O método de Sampaio-Stosic é obtido através do cálculo de influência que cada observação produz em todas as demais. Para captar a influência de cada observação os autores utilizam um estimador  $l_j$  (*leverage*) especificado na equação (15) abaixo:

$$l_j = \sqrt{\frac{\sum_{k=1; k \neq j}^k (\delta_{kj}^{\bullet} - \delta_k)^2}{K-1}} \quad (15)$$

---

<sup>29</sup> Ver Simar (2003, pp. 393-394). Um *outlier* é obtido por um DGP – *Data Generating Process* – diferente daquele que funciona para as demais observações ou então é uma informação muito rara de ocorrer para o DGP original. Os *outliers* podem ser também fruto de um simples erro de digitação.

O  $k$  é o conjunto de unidades de análise de 1 até  $K$ .<sup>30</sup> O “ $\delta_k$ ” denota o conjunto de índices de eficiência calculados para todas DMU’s  $\{\delta_{kj} \mid k = 1, \dots, K\}$ ; o  $\delta_{kj}^{\cdot}$  denota este conjunto de eficiência com a alavancagem, isto é, sem a DMU  $j$ ,  $\{\delta_{kj}^{\cdot} \mid k = 1, \dots, K \ k \neq j\}$ . Como observam os autores, a idéia básica é a de que se espera grande *leverage* das observações *outliers*. Se  $l_j$  destoar muito da média global tem-se a suspeita de ser um *outlier*. Quando a observação retirada está dentro da fronteira delimitada por  $P$  teremos  $(\delta_{kj}^{\cdot} - \delta_k) = 0$ , o que denota uma observação não-influente,  $l_j=0$ . No caso extremo de uma observação super-influente, sua retirada faz com que todas as outras subam para 1, assim  $\sum(\delta_{kj}^{\cdot} - \delta_k) = K-1$ , fazendo com que  $l_j=1$ . Portanto,  $0 \leq l_j \leq 1$ .

O teste de detecção de *outliers* Sampaio-Stosic será o primordialmente utilizado nesta dissertação. Maiores detalhes sobre sua construção estão presentes em Sampaio de Sousa e Stosic (2005, pp. 163-164).

O segundo teste de detecção de *outliers* empregado paralelamente foi proposto por Cazals et al. (2002), nomeado por CFS. O teste CFS usa um engenhoso princípio de “*trimming parameter*”. A idéia é construir fronteiras mais internas já que na presença de *outliers* a primeira fronteira a ser encontrada estará muito distante das outras informações. Os detalhes da construção do CFS estão no Anexo D.

Os métodos são empregados no capítulo 5 desta dissertação, os avanços da DEA e as interpretações que tais tratamentos permitiram acrescentar estão especificadas na análise. Por trás dos recentes métodos aqui expostos, existe uma série de outros trabalhos. A abordagem adotada procura focar a análise dos dados em primeiro plano, para tanto, os dois próximos capítulos da dissertação dialogam entre si:

---

<sup>30</sup> Aqui ocorre a mudança da notação utilizada por Sampaio-Stosic para evitar confusão entre a parte de cálculo de eficiência com a do cálculo dos *leverages*.

descreveremos primeiramente a base inicial e depois os resultados, utilizando os métodos explicados neste capítulo.

## 4. Fonte de Dados.

O emprego de dados educacionais sobre o desempenho e custo por aluno em Minas Gerais é um dos pontos relevantes desta dissertação. Os dados do SIMAVE permitem a abordagem de aspectos ligados à qualidade do ensino. A base do SICA, por sua vez, permite tratar de aspectos associados à eficiência.

As três subseções seguintes apresentam as fontes de dados utilizadas: 1) SIMAVE – Sistema Mineiro de Avaliação da Educação Pública; 2) SICA – Sistema Informacional Custo Aluno; e 3) Censo Educacional do MEC/IBGE. Essas três bases visam compor as variáveis de “outputs”, “inputs”, “dotação” e “background” presentes nos modelos desenvolvidos.

### 4.1. Dados do SIMAVE.

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 4ª e 8ª série da rede pública estadual, assim como os alunos do 3º 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.<sup>31</sup>

As informações obtidas com as provas são importantes para a avaliação da *qualidade* do ensino. Cada série a qual se aplica o exame possui uma matriz de referência que indica quais os níveis de conhecimento os alunos devem cumprir para o desempenho satisfatório em termos de aprendizado daquela etapa. 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

---

<sup>31</sup> Informações do SIMAVE no site: <http://www.simave.ufjf.br/home.asp>

comparação das notas entre os diversos anos do ciclo educacional. Assim sendo, podemos comparar a evolução das médias por série avaliada.

A Teoria de Resposta ao Item (TRI) elabora um método de interpretação das habilidades dos alunos por item específico de um caderno de questões. Cada item corresponde a uma habilidade particular da matéria (unidimensionalidade), não havendo interdependência entre as questões. Essa propriedade permite obter o resultado total das habilidades como o produtório das probabilidades de sucesso em cada item. Cada probabilidade de acerto indica uma habilidade específica do aluno, ou dos alunos. Se, por exemplo, um aluno acerta uma questão de aritmética com probabilidade total de acerto, isso denotará sua habilidade no quesito (que pode ser alta ou baixa, dependendo também da dificuldade da questão), para o aluno que erra com grande chance a habilidade será baixa.<sup>32</sup>

A habilidade é dada no eixo dos números reais e, ao contrário dos exames convencionais, não possui limite máximo nem mínimo. Normalmente adota-se a normal padronizada para a distribuição das habilidades. No caso do SIMAVE, as provas aplicadas às três séries utilizaram uma mesma escala de habilidades, onde a menor nota alcançada em matemática foi 75.47, de um aluno da 4ª Série e a maior foi 450.13, de um aluno da 3ª série. A comparação por séries pelo método TRI é obtida aplicando-se um número mínimo de questões comuns. As questões comuns servem de apoio para inferência do ganho de habilidade. Por exemplo, em relação à 4ª série, espera-se que um percentual bem mais elevado de alunos da 8ª série acertem as questões de aritmética ou álgebra simples, a proporção maior de acertos nesses itens permite verificar o ganho de habilidade.

---

<sup>32</sup> Soares (2005) verificou pela TRI e outras técnicas relacionadas ao exame de habilidades latentes, o número de questões mínimo para a discriminação das condições sociais dos alunos informadas no questionário do SIMAVE. Soares e Pereira (2002) verificaram as propriedades do TRI para a microrregião de Juiz de Fora, projeto do Laboratório de Medidas Educacionais (LAME) da universidade de Juiz de Fora que forneceu as condições de aplicação do SIMAVE ao estado. Segundo os autores, com algumas ressalvas, o teste aplicado desenvolve de maneira aceitável as propriedades do TRI. Por fim, Pereira (2006) analisa para as regiões Nordeste e Centro-Oeste do Brasil os fatores relacionados ao desempenho entre séries nas mesmas disciplinas aqui abordadas, Matemática e Português.

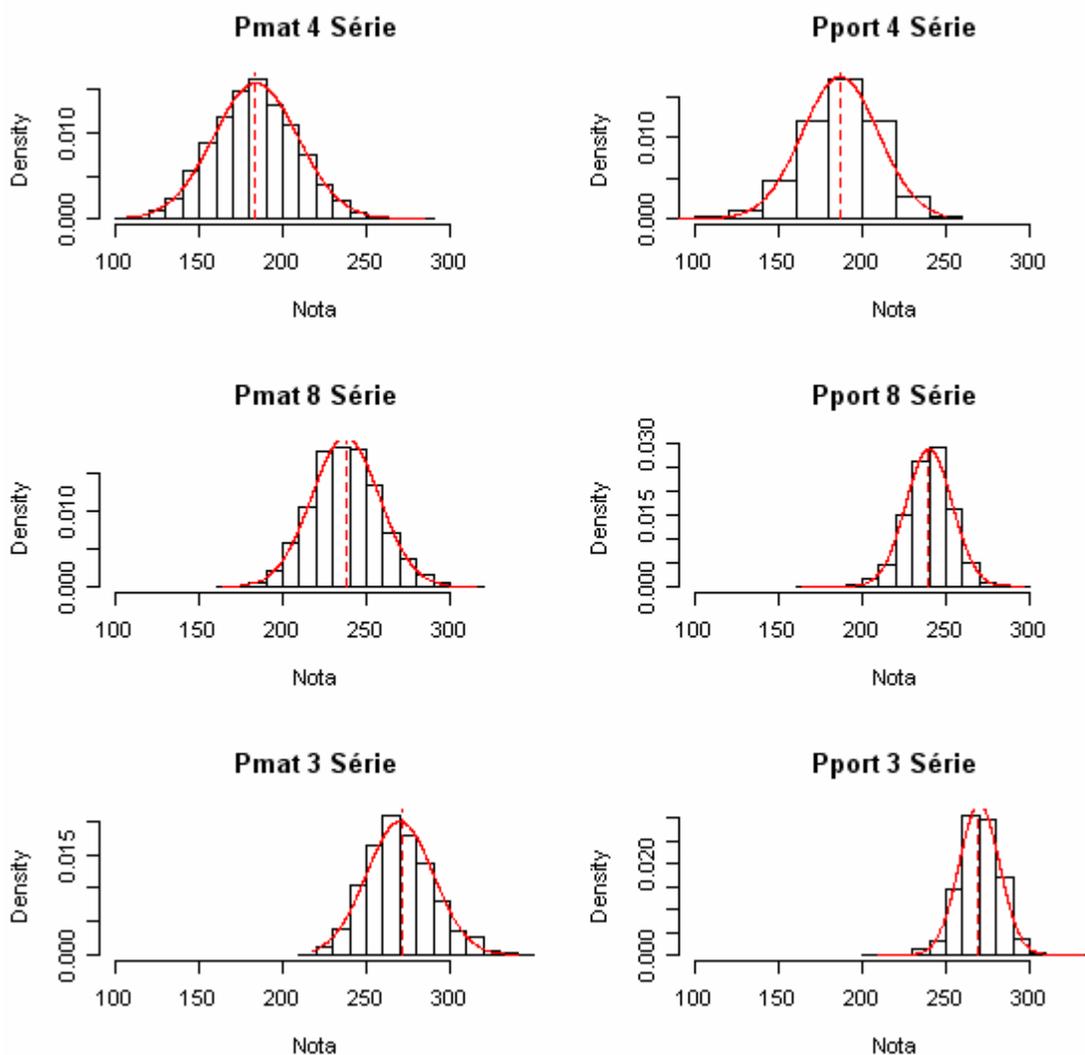
Os ganhos entre séries em termos de habilidades devem ser tomados com cautela. As propriedades de como se deve dar o avanço das notas não são muito bem estabelecidas. É normal supor que a habilidade média avance com o passar dos anos, no entanto, não é muito bem conhecido o processo como deve se dar esse avanço. Imagina-se que, à medida que os tópicos se tornam mais avançados, mais difícil é atingir um salto na progressão das notas, o que justifica a evolução observada nos dados do SIMAVE. No entanto, o valor observado na escala não reflete o esforço dos alunos para alcançar aquele resultado. A média de 183,73 de matemática na 4ª série pode indicar um cumprimento das metas de ensino maior do que a média 340.88 da 3ª série do ensino médio. A decisão sobre esta questão permanece um tanto normativa.

No gráfico a seguir apresentamos os histogramas comparados à distribuição normal das provas de matemática (2003) e português (2002) de cada uma das séries. Percebe-se que a distribuição de notas é bem próxima de uma distribuição normal e observa-se a evolução das médias por série, sendo o salto maior da 4ª para a 8ª e um pouco menor da 8ª para a 3ª série do ensino médio. A dispersão, medida pelo desvio padrão, diminui ao longo da evolução das séries.<sup>33</sup>

---

<sup>33</sup> Os dados originais das provas do SIMAVE estão desagregados por alunos. Como o objetivo é apresentar informações por escolas, os resultados são reportados pela média. Portanto, a distribuição dos histogramas apresentados é a distribuição de médias por escola e não dos alunos.

**Gráfico 4.1. Histograma e distribuição para as Notas de Matemática e Português.**



Fonte: SIMAVE, 2003 prova de matemática, *pmat*; 2002, prova de português, *pport*.

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.

Algumas escolas da base do SIMAVE possuíam informações para a prova de matemática, porém não possuíam os resultados da proficiência em português. Para escolas nessa situação foi realizada uma imputação do valor provável do resultado

da prova de português. Isso foi possível graças às outras informações disponíveis e, também, ao fato de a correlação entre a prova de matemática e português ser elevada (cerca de 0.68 para as três séries). Na 4ª série, 140 escolas se encontravam nessa situação (5% da amostra), na 8ª série, 86 (3%) e, na 3ª série, apenas 79 escolas tiveram o valor da proficiência estimados (novamente 5% das escolas).<sup>34</sup>

A análise da distribuição das notas revela o caráter normal e bastante simétrico deste tipo de informação. Não foram detectados *outliers* e as distribuições não requisitaram nenhuma transformação. A Tabela 4.1 a seguir resume as principais informações por disciplina e série.

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

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

Fonte: SIMAVE, 2003 prova de matemática, *pmat*; 2002, prova de português, *pport*.

<sup>34</sup> Método de imputação no ANEXO B.

Tão importantes quanto os exames de proficiência são as informações do questionário sócio-econômico das famílias aplicado pelo SIMAVE. A tabela abaixo resume a descrição dessas informações.

**Tabela 4.2. Variáveis do Questionário do SIMAVE.**

Variáveis para *background* (segundo estágio):

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

\* Para 8ª e 3ª séries a pergunta é apenas se leu livros.

Fonte: SIMAVE, 2003 questionário.

Os três quadros a seguir reúnem essas informações em forma de porcentagem para todas as escolas públicas estaduais de Minas Gerais e se apresentam nas três tabelas a seguir:

**Tabela 4.3. Resumo das informações do Questionário do SIMAVE 4ª Série.**

<b>Variável</b>	
<b>cor</b>	0.41
<b>esc_resp_fem</b>	0.31
<b>livros</b>	0.22
<b>computador</b>	0.13
<b>ler_livro_inf</b>	0.78
<b>ler_jornais</b>	0.44
<b>ntrab_fora</b>	0.84
<b>ntrab_domestico</b>	0.28
<b>devermat</b>	0.71
<b>nreprovado</b>	0.73

Fonte: SIMAVE 2003.

**Tabela 4.4. Resumo das informações do Questionário do SIMAVE 8ª Série.**

<b>Variável</b>	
<b>cor</b>	0.45
<b>esc_resp_fem</b>	0.19
<b>livros</b>	0.24
<b>computador</b>	0.11
<b>Ler_livro</b>	0.70
<b>Ler_jornais</b>	0.45
<b>trab_fora</b>	0.64
<b>trab_domestico</b>	0.26
<b>devermat</b>	0.55
<b>nreprovado</b>	0.67

Fonte: SIMAVE 2003.

**Tabela 4.5. Resumo das informações do Questionário do SIMAVE 3ª Série.**

<b>Variável</b>	
<b>cor</b>	0.51
<b>esc_resp_fem</b>	0.20
<b>livros</b>	0.26
<b>computador</b>	0.11
<b>Ler_livro</b>	0.71
<b>Ler_jornais</b>	0.62
<b>trab_fora</b>	0.47
<b>trab_domestico</b>	0.25
<b>devermat</b>	0.42
<b>nreprovado</b>	0.63
<b>noite</b>	0.69

Fonte: SIMAVE 2003.

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 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).<sup>35</sup>

#### **4.2. Dados do SICA.**

A segunda fonte de dados desta pesquisa é o SICA – Sistema Informacional Custo Aluno. O SICA fornece informações de gasto por aluno 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.<sup>36</sup> Como já foi mencionado, apesar do nome ‘custo’, as informações são de despesas na verdade. Tais dados são agrupados pelos sistemas periféricos que estão sob responsabilidade da Secretaria de Educação, Secretaria de Estado de Recursos Humanos e Administração e Secretaria de Estado da Fazenda. Não há, portanto, contabilização dos custos econômicos nem de receitas provenientes do setor privado.

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

---

<sup>35</sup> É o caso de refletir o quão positivos são estes indicadores: 16% das crianças de 9 a 11 anos ou 36% das crianças de 13 a 15 trabalhando fora de casa são percentuais ainda elevados, ainda quando não é possível distinguir ao certo a natureza deste trabalho.

<sup>36</sup> Sistema recém implementado pelo INEP: <http://www.siope.inep.gov.br/apresentacao.do>

peçoal administrativo e as despesas gerais. Os custos “fora” são aqueles que não estão direcionados ao custeio do peçoal 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. O ano dos dados custo-aluno utilizado é o último disponível, 2005.

Em relação às informações do banco de dados do SICA, foram precisos maiores cuidados no que diz respeito ao tratamento inicial. Para as informações de custo, foi necessário separar as informações das escolas especiais e de suplência da análise, por se caracterizarem como escolas de natureza diversa das que estão presentes na base inicial. Para se ter uma idéia, uma dessas escolas pertencia à primeira colocação em termos custos mais elevados, o valor era de R\$ 3719,00 por aluno-mês, 65 vezes maior do que a média da base sem *outliers*. Das 2469 escolas públicas estaduais da 4ª série, oito foram excluídas pelo motivo supracitado, duas por *missing* na variável de custos e apenas uma que apresentava custos excepcionalmente baixos R\$ 0,05 por aluno-mês. Na 8ª. série, são 2552 escolas no total, 16 excluídas por serem especiais, novamente duas por *missings* e uma de baixo custo. Para o 3ª série, os números são, respectivamente, 1453 escolas, 16 especiais, um *missing* e uma de custo baixo.

Após a retirada das escolas *outliers* no quesito custo, 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.<sup>37</sup> Abaixo temos a apresentação dos histogramas para estas variáveis. A curva em vermelho é a aproximação de uma normal, a curva azul é a densidade de Kernel-gaussiana. A média aparece na linha vertical pontilhada vermelha.

---

<sup>37</sup> No caso das regressões, observamos as transformações mais úteis para que um agrupamento dos dados seja mais representativo. Caba à variável *cdentro*, a transformação logarítmica que foi realizada. Mas para o método DEA, a transformação pode ser dispensada, primeiro porque não será testada a significância da variável e sim do índice final de eficiência e segundo porque o método de fronteiras não paramétricas se importa com o *ranking* dos insumos e produtos. Quando os dados possuem *outliers*, a transformação não altera a ordem relativa dos dados nem muito menos “conserta” a presença de *outliers*. Sobre isso ver as interessantes respostas aos comentários do *paper* seminal de Farrel (1957).

## Gráfico 4.2. Histogramas e Distribuições dos Custos Dentro por Série.

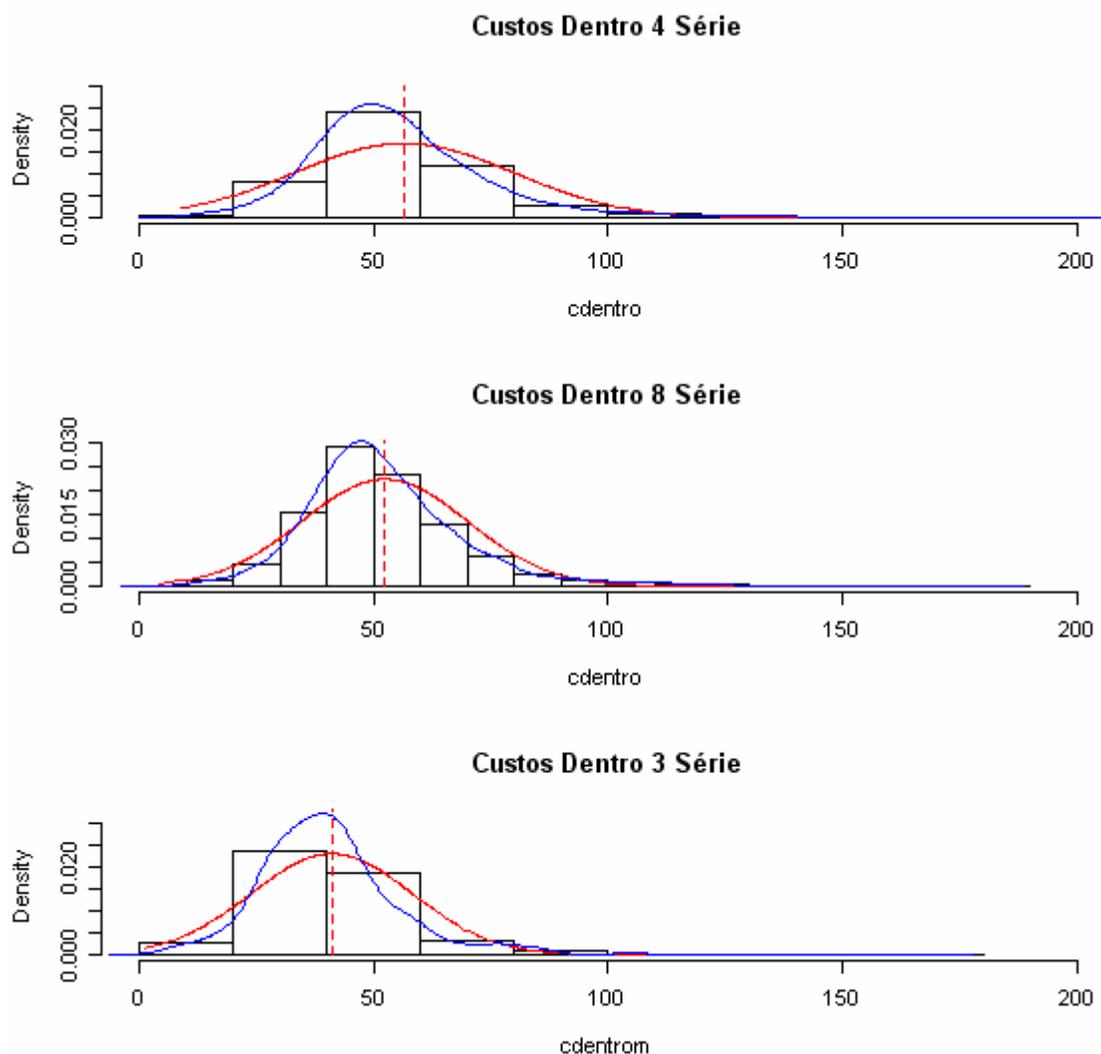
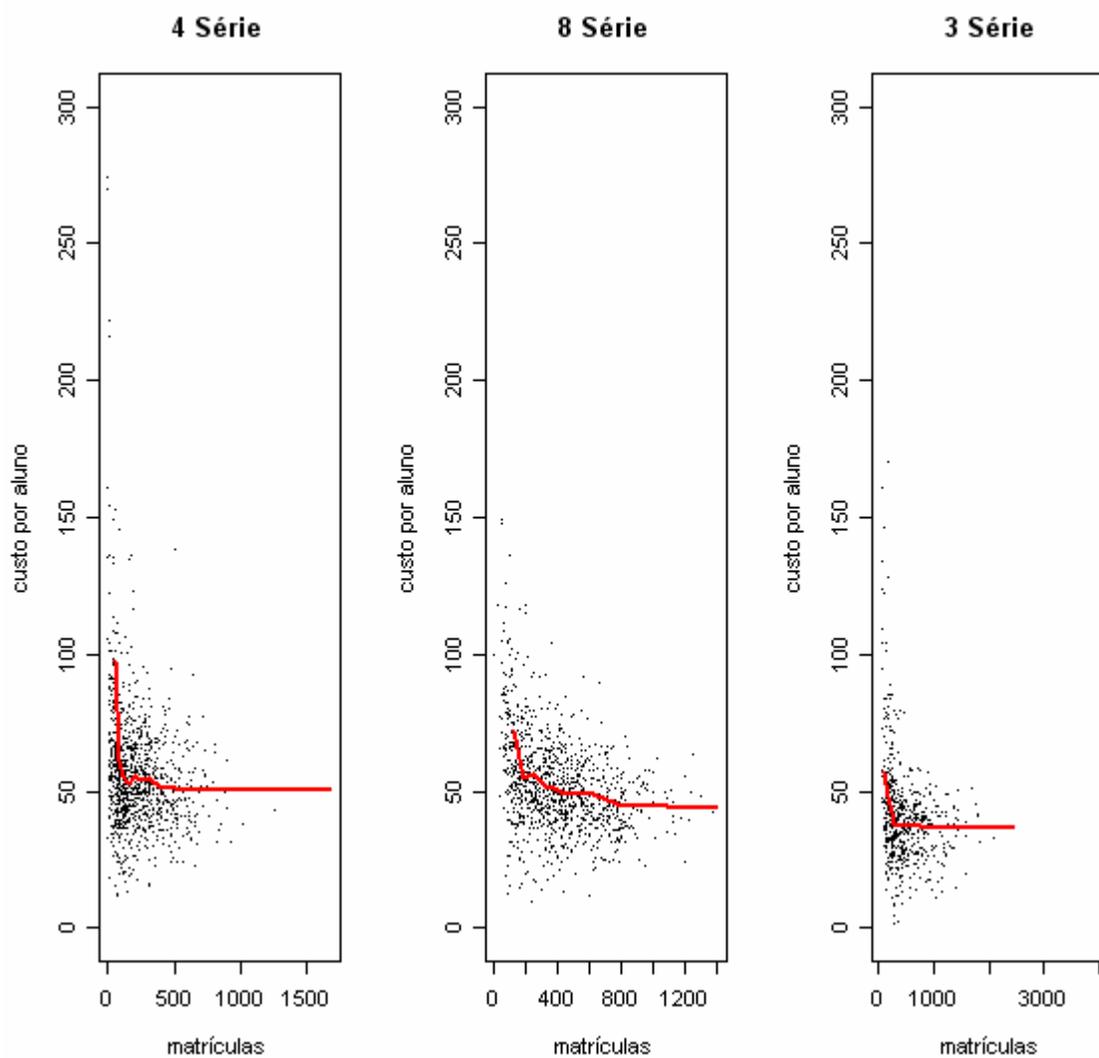


Figura 4.2 - Histogramas de Custos na Mesma escala para possibilitar a comparação

Fonte: SICA 2005.

Outra característica de interesse a ser mencionada para os dados de custo é o seu caráter decrescente em relação ao número de alunos. Observando o gráfico abaixo, onde são plotados o custo médio por aluno no eixo vertical e o número de matrículas por período no eixo horizontal, notam-se os ganhos de escala na manutenção de alunos em escolas que têm capacidade de absorver um número maior de matrículas. Além disso, observa-se que as curvas de custo alteram de patamar da 4ª série até a 3ª série, isso pode ser deduzido do aumento do número médio de alunos por escola a medida que se avança o nível de ensino.

**Gráfico 4.3. Curvas de Custo-Médio por Série.**



Fonte: SICA 2005.

Como *cdentro* incorpora basicamente as despesas variáveis (gasto com professores, diretores e materiais), temos uma aproximação de uma curva de custo variável médio.

### **4.3. Dados do Censo Escolar**

Por fim, a terceira fonte de dados é o Censo Escolar, realizado desde 1998 pelo INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), autarquia do Ministério da Educação (MEC). O Censo é uma base amplamente utilizada nos estudos sobre educação no Brasil. O questionário do Censo Escolar é respondido pelo diretor ou responsável de cada escola e engloba os níveis: Educação Infantil, Ensino Fundamental e Ensino Médio, nas modalidades: ensino regular, educação especial, educação de jovens e adultos e educação profissional de nível técnico. Contudo, este trabalho contempla apenas o nível Fundamental (1º e 2º ciclos) e Médio, na modalidade de ensino regular. O Censo Educacional utilizado é do ano de 2003.

Nesta dissertação, as variáveis do censo estão separadas em duas categorias: as agrupadas e as não agrupadas. A primeira categoria engloba os modelos DEA no primeiro estágio; são informações que se apresentaram mais úteis de maneira agrupada. 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.

**Tabela 4.6. Variáveis agrupadas e não-agrupadas do Censo Escolar.**

---

Variáveis Agrupadas (1º. Grupo):

<i>qdocente</i>	Número de professores com 3º Grau Licenciatura completa ("vdg155", "vdg156" e "vdg157"; "vdg165", "vdg166" e "vdg167"; "vdg175", "vdg176" e "vdg177") para 4ª 8ª série e 3º ano respectivamente).
<i>salas</i>	Número de salas de aula de toda escola - obtida pela soma de salas da variável "permanentes".
<i>Infra</i>	Variável para infra-estrutura presente na escola: Soma das dummies: <i>sani_den</i> - 1, Sanitário dentro da escola, <i>ener_pub</i> - 1 escola ligada a rede pública de energia elétrica; <i>agua_pub</i> - 1 escola com água da rede pública, <i>esg_pub</i> - 1 escola ligada ao esgotamento da rede pública, <i>lixo_col</i> - 1 destino do lixo coleta periódica.
<i>vcm</i>	Variável obtida por <i>Principal Component Analysis</i> para captar concomitantemente presença de aparelhos de "vídeo", "computadores" e recursos de "multimídia". Detalhes sobre a construção deste índice no <b>Anexo C</b> .

---

Variáveis não-agrupadas (2º. Grupo):

<i>urbana</i>	1- se escola se situa em área urbana, 0 caso contrário.
<i>matric_1a4</i>	numero de alunos matriculados no <b>primeiro</b> ciclo do fundamental (exclusiva p/ 4ª série).
<i>matric_5a8</i>	numero de alunos matriculados no <b>segundo</b> ciclo do fundamental (exclusiva p/ 8ª série).
<i>matric_médio</i>	numero de alunos matriculados no ensino médio (exclusiva p/ 3ª série do Ensino Médio).
<i>meso</i>	<i>dummy</i> para mesorregião do estado
<i>fnde</i>	1- escolas que recebem o Fundo Nacional de Desenvolvimento da Educação, 0 caso contrário.
<i>liv_ser4</i>	Séries que receberam livro didático para o ano do censo escolar - 4ª Série.
<i>liv_ser8</i>	Séries que receberam livro didático para o ano do censo escolar - 8ª Série.
<i>mere_esc</i>	1- escola oferece alimentação aos alunos.

---

O primeiro grupo resume as informações de maneira que o número de variáveis presentes na estimação da eficiência não seja demasiado. A convergência dos estimadores não-paramétricos da DEA é mais lenta do que a dos modelos paramétricos, a inclusão de muitos *inputs* e *outputs* atrapalharia a convergência dos estimadores tornando a análise menos eficiente em termos estatísticos (SIMAR e WILSON, 1998).

Dessa forma, construímos a variável *qdocentes* para agrupar em uma variável os professores com 3º grau: licenciatura completa ("vdg155"), sem licenciatura e com magistério ("vdg156") e 3º grau completo sem Licenciatura e sem Magistério ("vdg157").<sup>38</sup> As escolas estaduais de Minas Gerais se dividem basicamente entre

---

<sup>38</sup> Para a 8ª série essas variáveis de formação superior correspondem à VDG165-167, e para a 3ª série são as VDG175-177.

aquelas que possuem professores com licenciatura completa e as que possuem professores com 2º grau completo e magistério.

Observa-se, no histograma, a pequena quantidade de professores com formação superior atuando na 4ª série do fundamental. No entanto, esta média se eleva consideravelmente na 8ª série. Os dados refletem a prática pedagógica de separação de professores por turma. No primeiro ciclo do fundamental, cada turma possui apenas um professor responsável por todas as disciplinas, no segundo ciclo, cada matéria possui um professor, geralmente com licenciatura na cadeira. A especialização é ainda maior para a terceira série do Ensino Médio. No entanto, o número médio de professores com terceiro grau nesse nível de ensino cai um pouco em relação à etapa anterior, resultado de uma separação mais acentuada entre escolas sem e com professores qualificados.

**Gráfico 4.4. Histograma e distribuição de Kernel para *qdocentes*.**

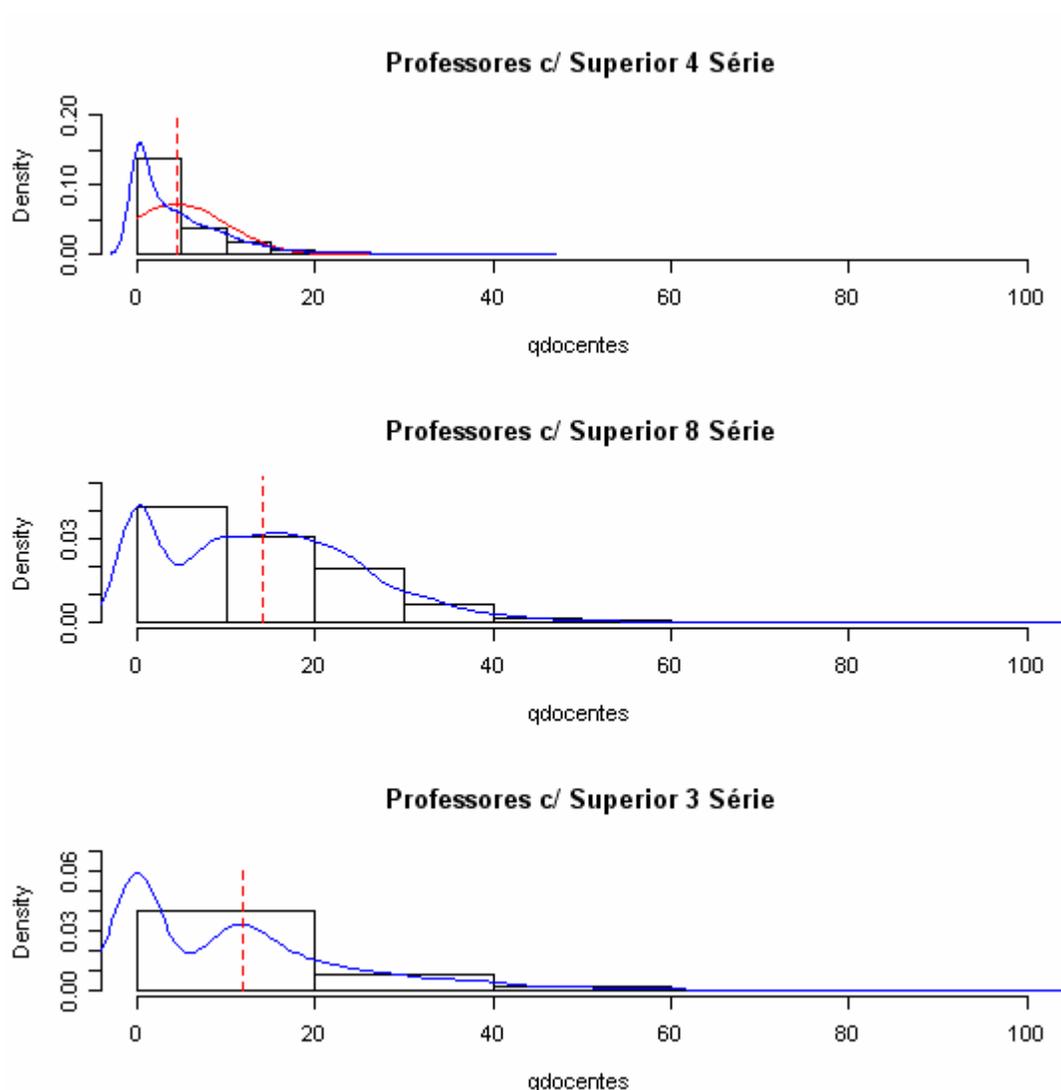


Figura 4.4 - Histogramas na Mesma escala

Fonte: Censo Escolar MEC/IBGE 2003. Variável *qdocente* construída, ver tabela 4.6.

A variável '*salas*' é também obtida de maneira simples. A partir do censo escolar, são selecionadas as informações sobre salas permanentes no estabelecimento, indicador que se constitui da soma de salas “dentro” e fora do “estabelecimento”. A distribuição de salas é a que se apresenta abaixo. Novamente, a curva vermelha é a distribuição normal e a azul de Kernel. Observa-se pequena evolução do número médio de salas por série.

#### Gráfico 4.5. Histograma para Numero de Salas.

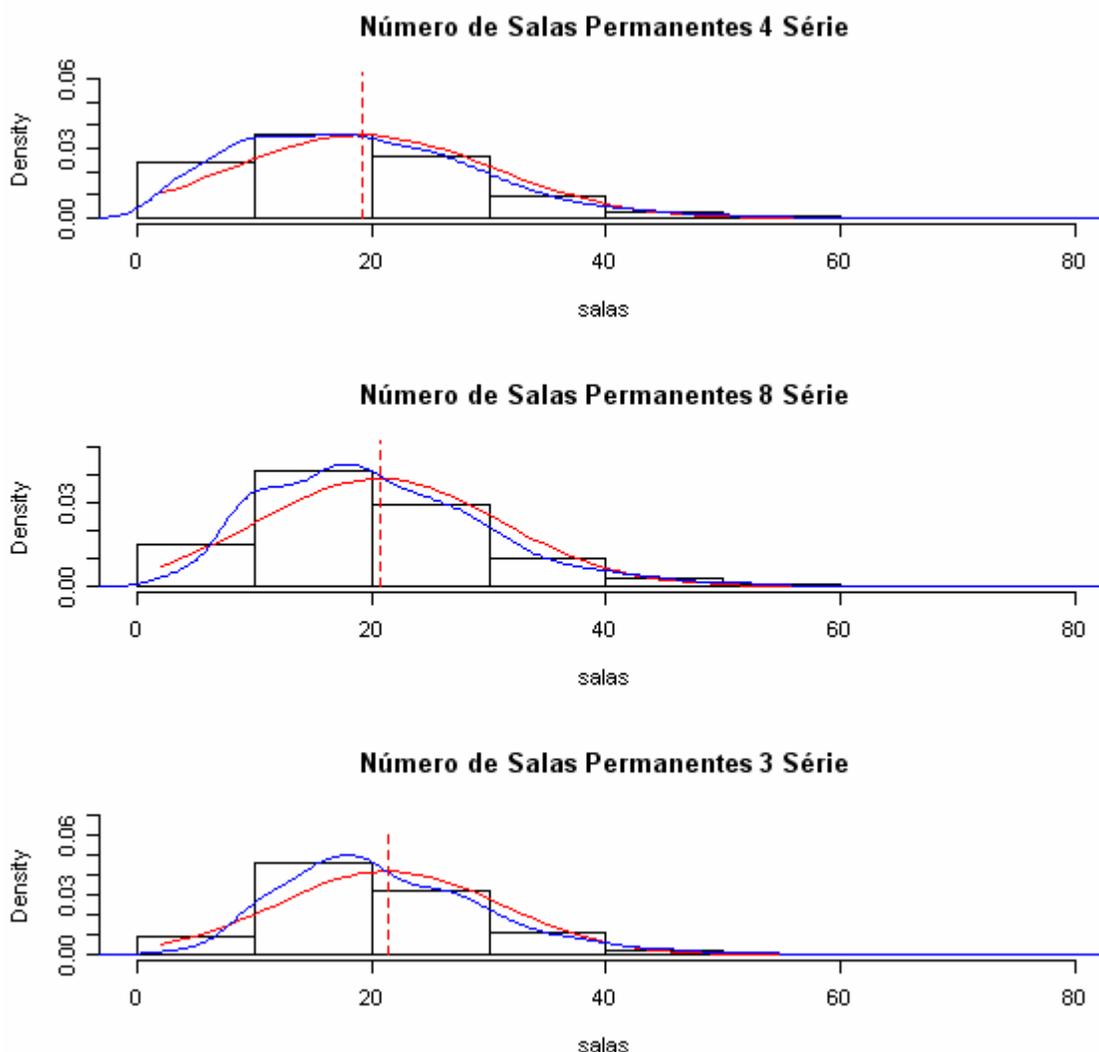


Figura 4.5 - Salas na Mesma escala. A média é ligeiramente maior para séries superiores

Fonte: Censo Escolar MEC/IBGE 2003. Variável 'Salas' construída, ver tabela 4.6.

Dentro da variável “condições de infra-estrutura das escolas (*Infra*)”, há cinco variáveis *dummy* somadas: *sani\_den* – observa se o sanitário está dentro da escola; *ener\_pub* – capta se a escola é ligada a 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. Os componentes da variável categórica *Infra* apresentavam elevada correlação entre si.

Abaixo temos o histograma, 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. A linha horizontal vermelha auxilia a observação de que para o terceiro ano, a proporção de escolas com todas as condições de infra-estrutura medidas é a maior, sendo que nenhuma escola possui índice zero neste último nível de ensino.

**Gráfico 4.6. Histograma para *Infra-estrutura*.**

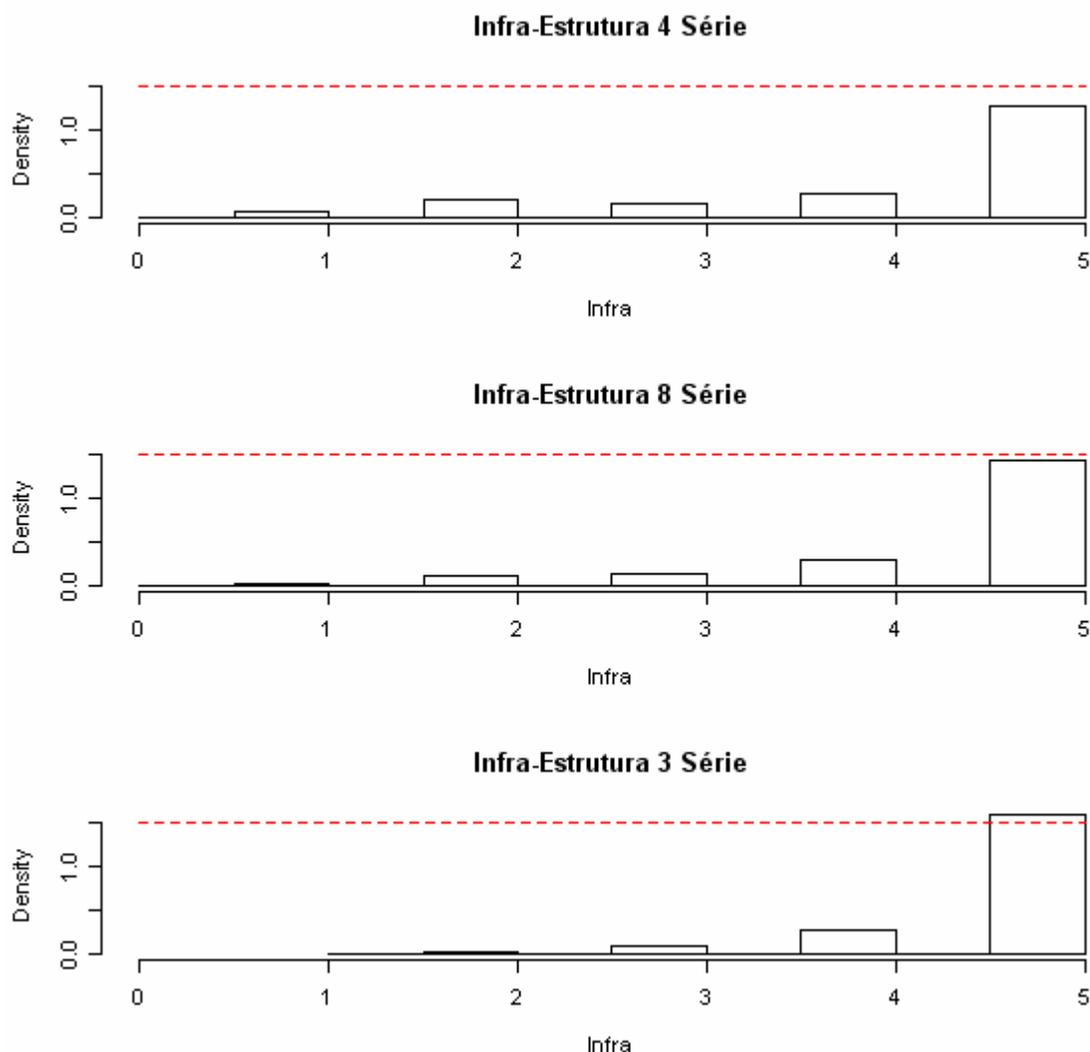


Figura 4.6 - Infraestrutura,  $y = 1.5$ . A soma da áreas é maior que 1 como ocorre nos dados de contagem

Fonte: Censo Escolar MEC/IBGE 2003. Variável '*Infra*' construída, ver tabela 4.6.

A variável *vcm* de *vídeo, computador e multimídia*, construída pelo método de PCA – *Principal Component Analysis*, combina variáveis que isoladas não se apresentavam muito significativas, possuíam pouca correlação com os resultados das provas, porém eram bastante correlacionadas entre si. "*Apvídeo*" 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. São agregados computadores (qualquer tipo) e impressoras. "*Mutimídia*" capta os recursos dentro da escola disponíveis aos alunos. Abarca a soma das variáveis originais: "*biblioteca*", *videotec* (videoteca), *saltvvid* (Sala de Tv e Vídeo), *sala\_lei* (Sala de Leitura) e *lab\_info* (laboratório de informática). O uso da *vcm* revelou-se importante, muito embora, à medida que se elevava o nível da escolaridade, perde a significância.

Como foi construída por *PCA* e deslocada para que não possuísse valores negativos (imposição do método), os histogramas de *VCM* não são comparáveis entre as séries. O que interessa no caso dessa variável é saber se as escolas têm melhores condições para os níveis de *VCM* mais altos. Importante observar a distribuição para cada série, a fim de observar o tamanho de sua dispersão. Constam as normais em vermelho para proceder tal verificação.

#### Gráfico 4.7. Histograma e distribuição Normal para *VCM*.

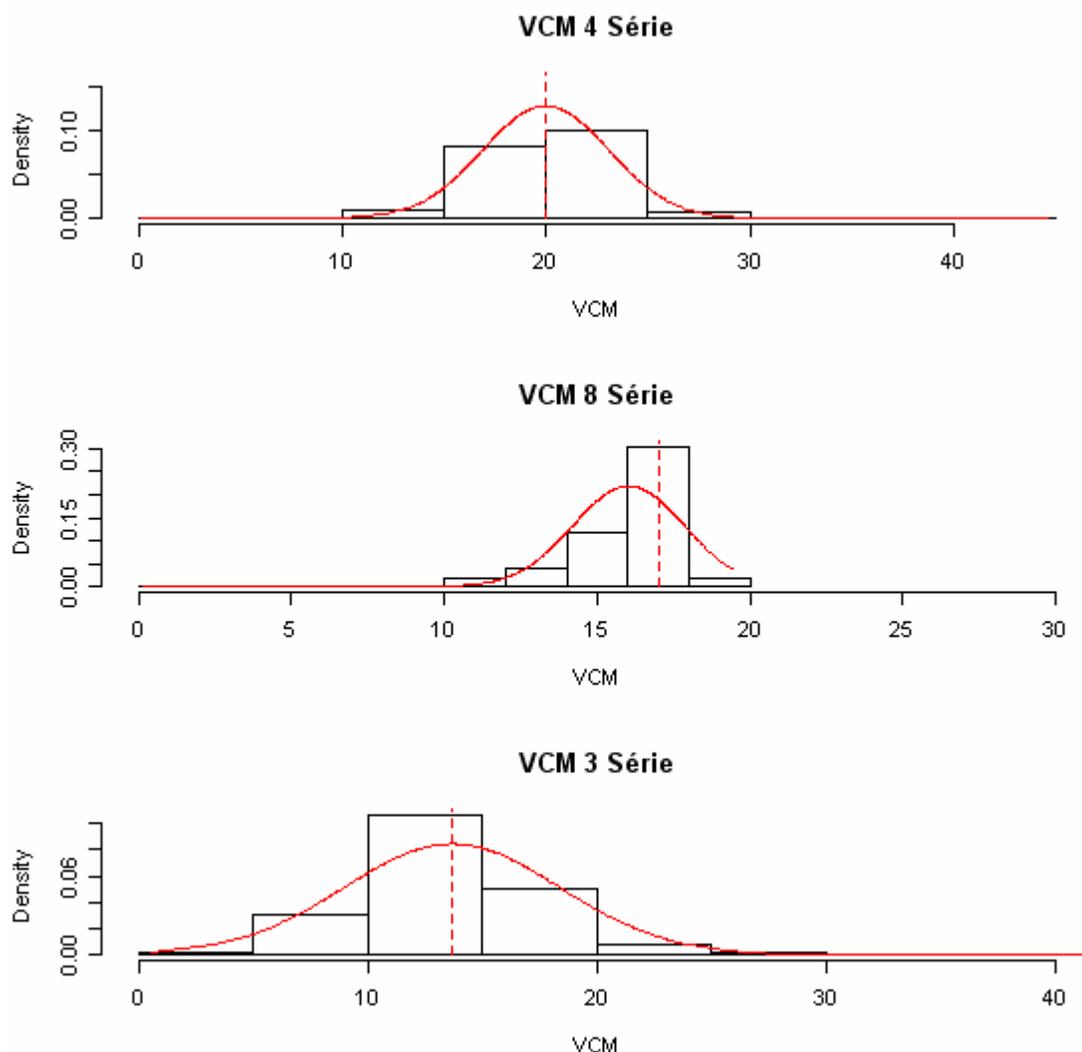


Figura 4.7 - Histograma VCM. Ressaltando que no caso dessa variável as escalas não são comparáveis.

Fonte: Censo Escolar MEC/IBGE 2003. Variável '*VCM*' construída por PCA, ver tabela 4.6.

Do grupo das desagregadas destaca-se *matrículas*, variável que será utilizada como indicador de *output*, já na primeira fase de estimação. 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 a 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 no histograma abaixo. As informações estão apresentadas por ciclo, matrículas de 1ª a 4ª série foi utilizada para a DEA 4ª série. A seguinte, matrículas 5ª a 8ª série, para a DEA 8ª série e *matric\_médio* é o total das 1ª, 2ª e 3ª séries do Ensino Médio, variável empregada para a análise da 3ª série.

**Gráfico 4.8. Histograma e distribuição de Kernel para *Matrículas por período*.**

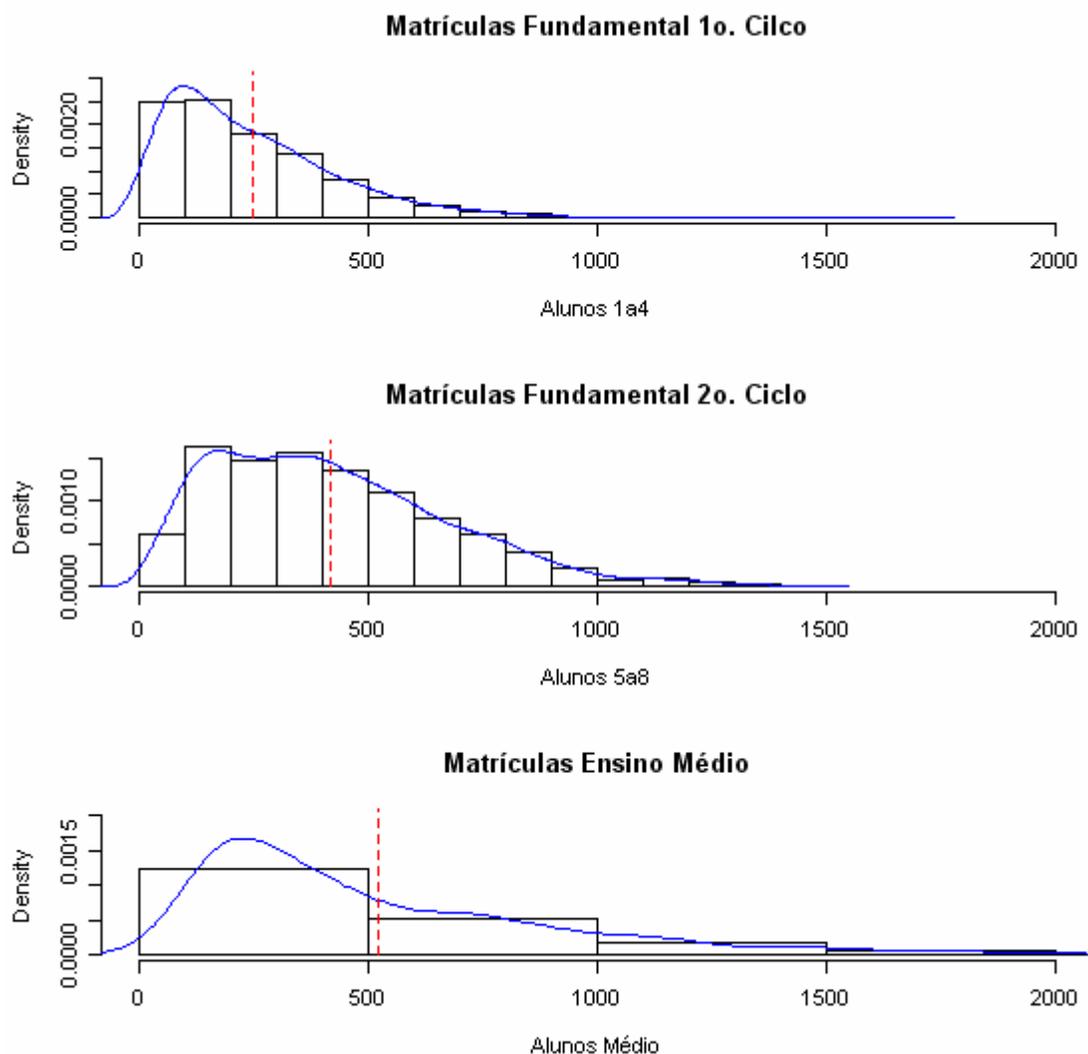


Figura 4.8 - Número de Alunos por Nível de Ensino. Há aumento da escala.

Fonte: Censo Escolar MEC/IBGE 2003. *matric\_1a4*, *matric\_1a5* e *matric\_médio*.

As demais variáveis do 2º. grupo não são abordadas detalhadamente nesta seção. Todas fazem parte do segundo estágio de estimação. Entretanto, apresentamos as tabelas que as resumem.

**Tabela 4.7. Resumo das informações do Censo escolar, variáveis não agrupadas 4ª Série.**

<b>Variável</b>	
urbana	0.79
4serie	0.79
fnde	0.81
liv4serie	0.25
liv8serie	0.34
merenda	0.96
TemTV	0.95
TemPC	0.58
TemMult	0.83
sani_den	0.94
ener_pub	0.98
agua_pub	0.83
esg_pub	0.71
lixo_col	0.77

Fonte: Censo Escolar MEC/IBGE 2003.

**Tabela 4.8. Resumo das informações do Censo escolar, variáveis não agrupadas 8ª Série.**

<b>Variável</b>	
urbana	0.89
8serie	0.79
fnde	0.79
liv4serie	0.20
liv8serie	0.42
Merenda	0.95
TemTV	0.98
TemPC	0.68
TemMult	0.89
sani_den	0.97
ener_pub	1.00
agua_pub	0.91
esg_pub	0.78
lixo_col	0.86

Fonte: Censo Escolar MEC/IBGE 2003.

**Tabela 4.9. Resumo das informações do Censo escolar, variáveis não agrupadas 3º Ano.**

Variável	
urbana	0.99
8serie	0.59
fnde	0.76
liv4serie	0.20
liv8serie	0.43
merenda	0.94
TemTV	0.99
TemPC	0.74
TemMult	0.94
sani_den	0.98
ener_pub	1.00
agua_pub	0.96
esg_pub	0.85
lixo_col	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 para auxílio de programas na escola), 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*).

Nos três bancos de dados disponíveis, estas variáveis apresentadas foram as escolhidas. Em qualquer método, a escolha final fica a critério do pesquisador. Isso ocorre em particular nos modelos DEA que são ditos discricionários. Ou seja, está a cargo dos avaliadores de eficiência escolher quais são os insumos mais importantes para se obter os produtos considerados relevantes.

Apesar da discricionariedade, existe uma orientação bastante utilizada pelos pesquisadores da DEA que é de incorporar apenas *inputs* e *outputs* que estão sob controle direto da unidade de decisão (a escola, no caso dessa pesquisa). 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. São os resultados de avaliação de eficiência e o delineamento de algumas propostas que estão nos dois próximos capítulos por vir, o resumo das variáveis escolhidas para o primeiro e segundo estágio estão nas tabelas 4.10 e 4.11. a seguir.

**Tabela 4.10. Resumo das variáveis de *outputs* e *inputs* escolhidas para o 1º estágio.**

Produtos	Insumos
– Matemática	– Custo Aluno
– Português	– Professores com curso Superior
– Matrículas por ciclo	– Número de Salas
	– Infra-estrutura
	– VCM

Fonte: SIMAVE (2002 e 2003). SICA (2005). Censo Escolar MEC/INEP 2003.

**Tabela 4.11. Resumo das variáveis *background* e *dotação* 2º estágio.**

Background	Dotação
– urbana	– urbana
– cor	– liv4serie
– série correta	– liv8serie
– escolaridade da mãe	– mere_esc
– livros	– apvideo
– internet	– comp_imp
– ler_livros	– multimid
– ler_jornais	– sani_den
– ntrab_fora	– ener_pub
– ntrab_domestico	– agua_pub
– devermat	– esg_pub
– nreprovado	– lixo_col
– noite	

Fonte: SIMAVE (2002 e 2003). SICA (2005). Censo Escolar MEC/INEP 2003.

Obs.: Cada uma das colunas compõe um tipo de regressão a ser visto no próximo capítulo. O regressando será o índice de eficiência (invertido) obtido no primeiro estágio.

## 5. Resultados.

Neste capítulo a análise de eficiência segue dois modelos. O primeiro trata do nível escola e implementa todas as técnicas já enunciadas. Os resultados são descritos para as doze mesoregiões de Minas Gerais. Este primeiro modelo vem acompanhado do método de dois estágios, onde estenderemos o alcance dos principais resultados encontrados relativos à eficiência da escola estadual mineira. O segundo modelo é a construção da eficiência a nível municipal, envolve mais variáveis do que as listadas no modelo de DEA por escolas.

### 5.1. Eficiência escolar mineira.

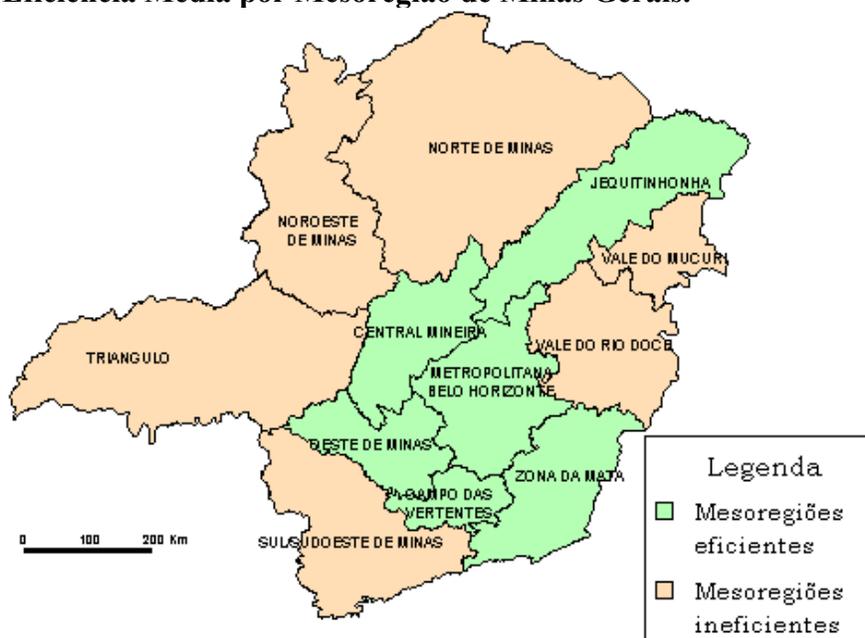
O índice de eficiência resume as variáveis e revela aspectos que seriam difíceis de serem captados por uma análise exaustiva da relação insumo produto solucionados um a um, ou através de uma simples interpretação de um conjunto extenso de informações. Assim, regiões que, a princípio, poderiam se sair 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 retratadas no mapa 5.1. A distribuição espacial da eficiência seguirá, para praticamente todas as séries, a configuração apresentada abaixo. As seis primeiras do *ranking* estão representadas na cor verde, as seis últimas na cor laranja. Observamos uma concentração das eficientes em torno da mesoregião Metropolitana, no fim do capítulo abordaremos melhor as explicações para esta configuração.<sup>39</sup>

---

<sup>39</sup> A mesoregião Sul/Sudoeste poderia ser incluída nas seis primeiras, em detrimento da mesoregião do Jequitinhonha, por exemplo, mas a motivação aqui é apresentar um padrão geral da eficiência e não uma distribuição da eficiência pelas mesos.

**Mapa 5.1. Eficiência Média por Mesoregião de Minas Gerais.**



Fonte: Mapa Geopolítico de Minas Gerais IGA/CETEC 1994. Modificado.

### 5.1.1. Eficiência 4ª série.

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

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

Código	Mesoregião	#	Produtos			Insumos				
			<i>Pmate</i>	<i>Pport</i>	<i>Alunos</i>	<i>Custo</i>	<i>Qdoc</i>	<i>Salas</i>	<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
<b>310013</b>	<b>Minas Gerais</b>	<b>2458</b>	<b>183.73</b>	<b>186.72</b>	<b>247.64</b>	<b>58.41</b>	<b>4.36</b>	<b>19.21</b>	<b>4.24</b>	<b>19.98</b>

Fonte: Elaboração própria.

Constatamos que, via de regra, em Minas Gerais, os custos por alunos aumentam à medida que se desloca para o sul do estado. A Região Metropolitana é uma das que possui maior média para os produtos, em contrapartida, notamos que o nível de insumos utilizados nessa meso é também um dos mais elevados. A menor média para as notas se verifica na região Norte do estado, é também a região que apresenta segunda menor média no número de matrículas, além disso, o nível de insumos no Norte 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, muitos insumos e pouco produto denota ineficiência, muito produto para pouco insumo, evidencia evento improvável, grande chance da observação ser *outlier*. Na análise para a 4ª série convêm destacarmos a mesoregião do Jequitinhonha, tal região é caracterizada como uma das mais pobres de Minas, entretanto, isso não a impede de se sair bem no *ranking* de eficiência.

**Tabela 5.2. Ranking por DEA-*bootstraped* (DEA\_c\*) para a 4ª série:**

Código	Mesorregião	#	DEA_c*	Rank1
310009	Oeste de Minas	90	0.654	1
310007	Metropolitana	527	0.650	2
310011	Campo das Vertentes	61	0.650	3
310012	Zona da Mata	293	0.647	4
310006	Central Mineira	55	0.646	5
310003	Jequitinhonha	191	0.641	6
310001	Noroeste de Minas	54	0.629	7
310010	Sul/Sudoeste de Minas	183	0.629	8
310008	Vale do Rio Doce	303	0.623	9
310002	Norte de Minas	411	0.623	10
310005	Triângulo Mineiro	192	0.621	11
310004	Vale do Mucuri	98	0.607	12
<b>310013</b>	<b>Minas Gerais</b>	<b>2458</b>	<b>0.635</b>	

Fonte: Elaboração própria.

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 Triângulo Mineiro é

uma região que vêm alcançando grande desenvolvimento social e econômico e, de fato, algumas escolas dos municípios populosos de Uberlândia e Uberaba denotam isso, no entanto, as melhores escolas eficientes da região ainda não conseguem contrapor o relativamente fraco desempenho de produto das demais escolas o que explica o fraco desempenho, na média, dessa meso, tendo em vista um dos mais altos níveis de insumos alocados do estado.

O *ranking* de eficiência é sensível ao tipo de índice adotado (em relação às DEA's de escala, não muito). No que se refere ao método de construção, pode variar de forma significativa. Vimos que o índice DEA clássico está sujeito à uma série de influências importantes que distorcem sua interpretação (WILSON 1995; SIMAR 2003; SAMPAIO de SOUSA e STOSIC 2005). 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). O índice Sampaio-Stosic detecta *outliers* e confecciona novos índices após descartar as observações super-influente. Em comparação com uma lista de escolas de qualidade fornecida pela Secretaria de Educação do Estado de Minas Gerais (SEE/MG), o Sampaio-Stosic foi o índice corrigido que apresentou maior aderência<sup>40</sup>, sendo um método muito útil para, além de apontar *outliers*, identificar as escolas que servem de referência.

---

<sup>40</sup> Anexo D.

**Tabela 5.3. Ranking índice Sampaio-Stosic (DEA\_ss) para 4ª série:**

Código	Mesorregião	#	DEA_ss	Rank2
310007	Metropolitana	527	0.767	1
310002	Norte de Minas	411	0.760	2
310011	Campo das Vertentes	61	0.747	3
310009	Oeste de Minas	90	0.757	4
310003	Jequitinhonha	191	0.755	5
310012	Zona da Mata	293	0.747	6
310006	Central Mineira	55	0.744	7
310008	Vale do Rio Doce	303	0.734	8
310010	Sul/Sudoeste de Minas	183	0.732	9
310001	Noroeste de Minas	54	0.731	10
310005	Triângulo Mineiro	192	0.724	11
310004	Vale do Mucuri	98	0.701	12
<b>310013</b>	<b>Minas Gerais</b>	<b>2458</b>	<b>0.748</b>	

Fonte: Elaboração própria.

Após a exclusão dos *outliers*, a mesoregião que mais se beneficiou foi a Norte de Minas, passando para o primeiro grupo. A meso que deixou de configurar entre as seis primeiras foi a Central Mineira, caindo duas posições. O Norte era a segunda meso que mais possuía escolas declaradas *outliers*, 9 sobre o total de 35 (a primeira é a metropolitana com 14 *outliers*). As escolas *outliers* prejudicavam a comparação da região como um todo.<sup>41</sup>

Outro *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 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. O *ranking* do quociente serve mais para relativizar as médias de eficiência por região, diferentemente dos outros dois, não é um *ranking* de eficiência no sentido estrito. Observe que a Zona da Mata é a que mais perde posições em relação ao *rank1*. Dado o número total de escolas na Zona da Mata, esperávamos bem mais escolas declaradas eficientes. O contrário ocorre com a mesoregião Norte

<sup>41</sup> 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.

de Minas, a região possui uma razão elevada de escolas eficientes sobre o total de escolas. O Vale do Rio Doce, do Mucuri e o Triângulo Mineiro, também melhoram um pouco nesse aspecto.

**Tabela 5.4. Ranking do quociente eficientes/total de escolas 4<sup>a</sup> série:**

<b>Código</b>	<b>Mesorregião</b>	<b>#</b>	<b>quocient</b>	<b>Rank3</b>
310011	Campo das Vertentes	61	0.082	1
310006	Central Mineira	55	0.055	2
310002	Norte de Minas	411	0.054	3
310007	Metropolitana	527	0.047	4
310008	Vale do Rio Doce	303	0.040	5
310009	Oeste de Minas	90	0.033	6
310010	Sul/Sudoeste de Minas	183	0.027	7
310005	Triângulo Mineiro	192	0.021	8
310004	Vale do Mucuri	98	0.020	9
310001	Noroeste de Minas	54	0.019	10
310012	Zona da Mata	293	0.017	11
310003	Jequitinhonha	191	0.016	12
<b>310013</b>	<b>Minas Gerais</b>	<b>2458</b>		

Fonte: Elaboração própria.

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. O DEA-C é rejeitado apenas para a 3<sup>a</sup> série do Ensino Médio.<sup>42</sup> No entanto, a construção dos três índices é importante para captar as escolas por grupo de retornos de escala (como mostrado na figura 3.1). Por meio daquele procedimento, detectamos a distribuição de escalas de rendimento por mesoregião de Minas Gerais. Na tabela abaixo, cada uma das linhas soma 100%.

<sup>42</sup> Ver os testes para a escala no ANEXO C.

**Tabela 5.5. Distribuição das escalas de rendimento por Mesoregião:**

<b>Código</b>	<b>Mesorregião</b>	<b>#</b>	<b>Constante</b>	<b>Decrecente</b>	<b>Crescente</b>
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%
<b>310013</b>	<b>Minas Gerais</b>	<b>2458</b>	<b>3.13%</b>	<b>94.34%</b>	<b>2.52%</b>

Fonte: 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 tendência maior a apresentar proporções mais elevadas de escolas com rendimentos **constant**es e **cre**scentes.

A Região Metropolitana, por ser a mais populosa e apresentar as maiores escolas, possui 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.<sup>43</sup> Em termos de política, isso confirma que insumos alocados nas regiões mais pobres obtêm maior aproveitamento.

<sup>43</sup> A rigor, não faz sentido definir escala para escolas não-eficientes, mas a construção da fronteira permite identificar a escala ideal em que deveriam estar atuando as observações ineficientes. (Wilson 2005a, pp. 17).

A Tabela abaixo resume todas as informações até aqui apresentadas e será utilizada, de agora em diante, para as demais séries apresentadas nesse capítulo. As informações da tabela 5.5 estão resumidas pelo número de escolas **eficientes** declaradas como rendimentos decrescentes de escala (*Descresc*).

**Tabela 5.6. Ranking das mesorregiões para a 4<sup>a</sup> série:**

<b>Código</b>	<b>Mesorregião</b>	<b>#</b>	<b>DEA_c*</b>	<b>Rank1</b>	<b>DEA_ss</b>	<b>Rank2</b>	<b>quocient</b>	<b>Rank3</b>	<b>Decresc</b>
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
<b>310013</b>	<b>Minas Gerais</b>	<b>2458</b>	<b>0.635</b>		<b>0.748</b>		<b>0.037</b>		<b>28</b>

Fonte: Elaboração própria.

O principal objetivo na identificação de escolas eficientes é apresentar a comparação entre os grupos. As tabelas abaixo fornecem a comparação entre os produtos e insumos empregados. O critério para a separação entre eficientes e ineficientes é flexível. No caso da representação abaixo, utilizamos os 30% melhores colocados no ranking (736 escolas).<sup>44</sup>

**Tabela 5.7. Resumo das Variáveis da Função para Escolas Eficientes 4 série:**

	<b>Média</b>	<b>Desvio</b>	<b>Min</b>	<b>Max</b>
<b>Matemática</b>	194.14	25.18	106.72	277.83
<b>Português</b>	198.92	19.47	126.26	249.86
<b>Matrículas 4 série</b>	268.35	207.76	6.00	1170.00
<b>Matrículas Escola</b>	1200.68	856.49	12.00	4422.00
<b>Custo Dentro</b>	54.39	24.13	11.39	304.10
<b>Docentes com superior</b>	3.55	5.00	0.00	39.00
<b>Salas</b>	17.87	10.17	2.00	56.00
<b>Infraestrutura</b>	4.10	1.30	0.00	5.00
<b>VCM</b>	17.99	2.54	8.21	24.77

Fonte: SIMAVE 2002, 2003; Censo Escolar MEC/IBGE 2003; SICA 2005.

**Tabela 5.8. Resumo das Variáveis da Função para Escolas NÃO Eficientes 4 série:**

	<b>Média</b>	<b>Desvio</b>	<b>Min</b>	<b>Max</b>
<b>Matemática</b>	179.28	23.90	113.44	284.93
<b>Português</b>	181.50	22.21	53.25	244.01
<b>Matrículas 4 série</b>	238.79	186.26	5.00	1675.00
<b>Matrículas Escola</b>	1005.42	747.60	10.00	6126.00
<b>Custo Dentro</b>	60.13	23.42	9.20	329.33
<b>Docentes com superior</b>	4.71	5.53	0.00	44.00
<b>Salas</b>	19.79	11.53	2.00	136.00
<b>Infraestrutura</b>	4.30	1.16	0.00	5.00
<b>VCM</b>	20.83	2.92	0.15	44.64

Fonte: SIMAVE 2002, 2003; Censo Escolar MEC/IBGE 2003; SICA 2005.

<sup>44</sup> No uso da DEA-clássica a separação entre eficientes e ineficientes comumente se dá entre as observações com índice 1 (escolas “sobre” a fronteira de eficiência) e as de índice inferior (interior da fronteira). Com o uso do *bootstrap*, esse tipo de separação pode ser mais flexível, já que nos referimos agora à eficiência em média provável e não apenas a **um** índice exato.

Observamos que as escolas eficientes possuem melhor desempenho nas notas e um custo por aluno ligeiramente menor. Possuem, também, a mesma infra-estrutura e quase o mesmo número de salas e um índice menor para o aparelhamento escolar (VCM), ou seja, o grupo das 30% mais eficientes não são todas escolas bem providas em equipamentos de tv vídeo e parabólica.

As regressões (1) e (2) na tabela 5.9 a seguir são estimações das normais truncadas obtidas por máxima verossimilhança (MV), já que as variáveis regredidas possuem o limite máximo na eficiência igual a 1. As duas primeiras regressões seguem o algoritmo proposto em Simar e Wilson (2003) e utilizado também em Afonso e Aubyn (2005).<sup>45</sup>

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 urbana-rural. Estas duas primeiras regressões servem de “guia” para as quatro demais, pois são mais rigorosas na construção dos intervalos dos testes de hipóteses e rejeitam a hipótese nula com menor frequência.

As regressões (3) e (4), na tabela 5.9, são regressões normais-truncadas por máxima verossimilhança, no entanto, sem seguir os passos do algoritmo Simar e Wilson (2003). Nas regressões (5) e (6) temos o modelo tobit na forma usual de regressão de dois-estágios empregada (McCARTY e YAISAWARGN, 1993). A diferença na regressão (6) é a utilização da eficiência não-corrigida no regressando.<sup>46</sup> Na quarta série, com exceção de ‘*ntrab\_fora*’, todas as variáveis significativas apresentam a mesma direção das regressões (1) e (2).

---

<sup>45</sup> ANEXO E fornece detalhes do algoritmo.

<sup>46</sup> Mesmo usando uma regressão em sua versão mais *naivë*, os dados parecem apontar para a direção esperada.

É preciso observar que o índice precisou ser alterado de 1 a  $\infty$  para adaptar à função programada<sup>47</sup>. Os sinais das variáveis significativas se apresentam na direção esperada. A surpresa é ‘urbana’, o coeficiente das regressões (1), (3), (5) e (6) é positivo, indicando que se a escola é urbana, essa tende a ser uma escola menos eficiente em probabilidade.<sup>48</sup> 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.<sup>49</sup> A última interpretação ganha reforço, porque, ao controlarmos as regressões pelas condições de infra-estrutura, menores no meio rural (regressões 2 e 4), o coeficiente negativo passa a 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. A análise de eficiência indica que, das escolas declaradas eficientes, há maior proporção de mães de alunos com mais de 8 anos de estudo. As escolas eficientes absorvem filhos de pais com maior escolaridade, ou, pais mais instruídos optam, com maior frequência, matricular seus filhos em escolas eficientes.

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

---

<sup>47</sup> O índice utilizado nas regressões é o inverso ( $1/\delta_i$ ), vai do mais eficiente 1, até o menos eficiente, maior que 1. A interpretação é: se o coeficiente é positivo se afasta da eficiência, se negativo, aproxima-se.

<sup>48</sup> Se nos deslocarmos do meio rural para o urbano, maior será a proporção de escolas ineficientes nesse meio em relação às rurais.

<sup>49</sup> É verdade que o tipo de escala importa nessa análise. Se maior parte das escolas urbanas se situam na parte de rendimentos decrescentes de escala então, usando o DEA-V, teríamos mais urbanas eficientes, mas foi para estabelecer essa questão que realizamos o teste para escala presente no ANEXO D.

como significativa a 1% em todas as outras regressões, como se verifica, também, para a 3ª série. Não podemos dizer o mesmo para o caso dos alunos que não trabalham fora (*ntrab\_fora*), apesar de nas demais regressões o coeficiente se apresentar significativo e na direção esperada pela intuição, as informações não se confirmam na primeira regressão.

A indicação sobre se os alunos na sala de aula lêem, '*ler\_livro\_inf*', é também um interessante ponto afirmativo de eficiência. As escolas eficientes podem ser aquelas que conseguem estimular mais a leitura, ou então, são aquelas que atraem alunos com maior interesse por leitura. Em igual maneira para a realização do dever de matemática (*devermat*).

O desafio para o entendimento está no percentual de reprovações para escolas eficientes. O coeficiente de '*nreprovado*' é positivo, o que indica que as escolas que menos reprovam são as ineficientes. Isso pode ocorrer devido à uma diferenciação entre as escolas públicas em Minas Gerais. A adoção da **progressão continuada**, que avança os alunos automaticamente para a série seguinte pode ser um dos fatores que explicam porque as escolas com o menor índice de reprovância são menos eficientes. Dentro da **progressão seriada**, os alunos repentes possuem chance de aprender o conteúdo daquela série e se desempenharem melhor no exame de proficiência.

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 considerável de escolas da 4ª série possuem 10 ou mais televisores (120 escolas).<sup>50</sup> O resultado aponta é que as escolas

---

<sup>50</sup> 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

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 contribui para o desempenho escolar.

Todavia, é interessante observar que é o contrário para o caso dos computadores (*comp\_imp*). Na 4ª série são 1024 escolas sem computador, destas, 82 (8%) foram declaradas eficientes, das 629 com 10 ou mais computadores 13 (18,7%) eram eficientes.<sup>51</sup> Os computadores nas escolas contribuem para sua eficiência. A explicação que se delineia pode ser a seguinte: a técnica pedagógica vigente consegue utilizar o computador como uma ferramenta na melhoria do ensino, enquanto que os demais recursos parecem não estar contribuindo tanto, o que pode denotar um sub-uso ou uma utilização não proveitosa para as notas dos alunos em português e matemática.<sup>52</sup>

As outras variáveis de infra-estrutura são menos significativas, porém, não indicam fatores tão depreciativos. A forma de coleta de água da escola não necessita ser necessariamente pública, assim como há diversas maneiras de coleta do lixo. O indicador de presença ou ausência do quesito de infra-estrutura é desenvolvido no ANEXO E.

---

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*.

<sup>51</sup> Para a 8ª e 3ª série essa comparação é semelhante.

<sup>52</sup> É sempre importante lembrar as limitações da análise: é demonstrado que a televisão pode ser um recurso útil no ensino (Moran 2002), no entanto, na presença dos poucos fatores analisados e de provas apenas de português e matemática, o equipamento parece não ter efeito. Mas há, logicamente, a limitação da análise que observa a questão do ensino sob o ângulo específico da eficiência. A proposta construtiva que pode surgir é a de se perguntar como os fatores não significativos de dotação de equipamentos podem contribuir para incrementar os resultados.

**Tabela 5.9. Resultados das Regressões dois estágios 4ª série.**

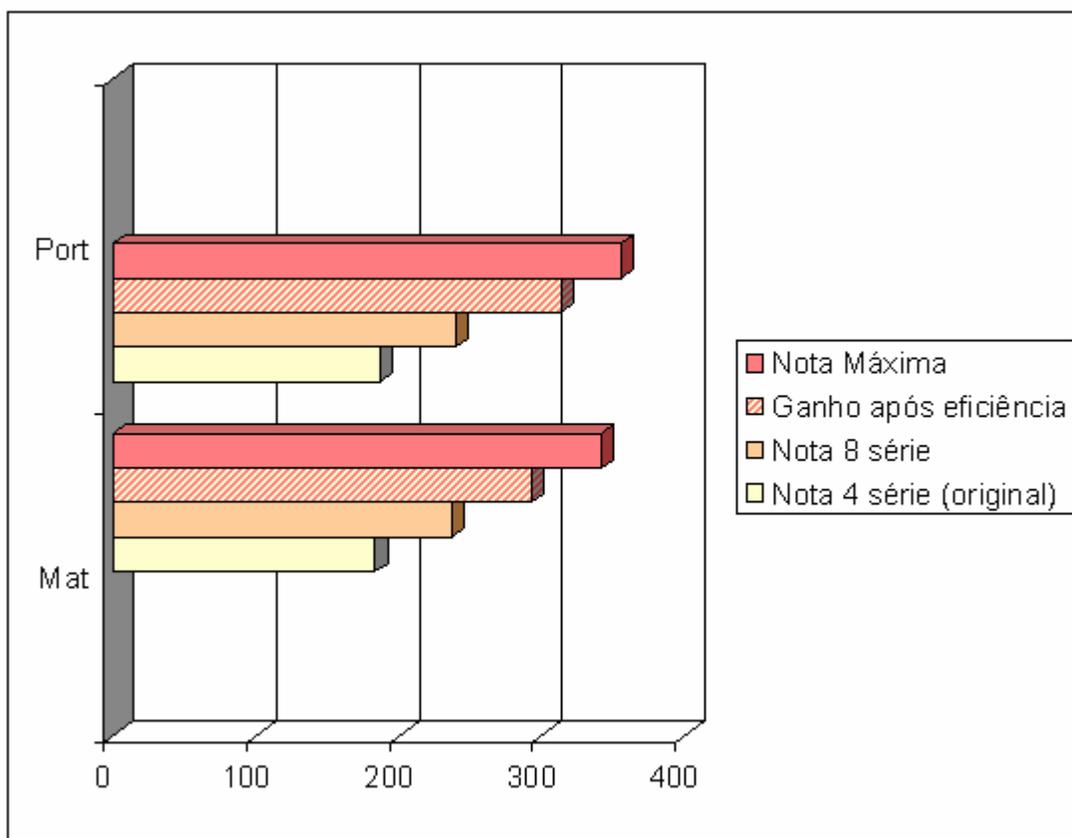
	Regressão (1)	Regressão (2)	Regressão (3)	Regressão (4)	Regressão (5) (tobit1)	Regressão (6) (tobit2)
constante	1.238*** (0.05)	0.253** (0.12)	1.847*** (0.05)	1.554 (0.03)	1.841*** (0.05)	1.584*** (0.05)
urbana	0.136*** (0.05)	-0.216*** (0.01)	0.084*** (0.01)	-0.075*** (0.02)	0.081*** (0.01)	0.164*** (0.01)
fnde	0.071 (0.09)		0.035*** (0.01)		0.033*** (0.01)	0.041*** (0.01)
mere_esc	0.001 (0.10)		0.002 (0.02)		0.002 (0.02)	0.074*** (0.02)
cor	0.025 (0.13)		-0.013 (0.03)		-0.012 (0.03)	-0.018 (0.03)
esc_resp_fem	-0.385** (0.16)		-0.145*** (0.04)		-0.139*** (0.03)	-0.100*** (0.03)
livros	-0.189 (0.18)		-0.127*** (0.04)		-0.121*** (0.04)	-0.125*** (0.04)
computador	-0.143 (0.12)		-0.088* (0.05)		-0.083* (0.04)	-0.170*** (0.05)
ler_livro_inf	-0.041 (0.10)		-0.104*** (0.03)		-0.100*** (0.03)	-0.090 (0.03)
ler_jornais	-0.088 (0.13)		-0.056** (0.03)		-0.054** (0.03)	-0.052** (0.03)
ntrab_fora	0.092 (0.12)		-0.135*** (0.04)		-0.130*** (0.04)	-0.122*** (0.04)
ntrab_domestico	0.053 (0.11)		-0.006 (0.03)		-0.005 (0.03)	-0.005 (0.03)
devermat	-0.087 (0.11)		-0.102*** (0.03)		-0.098*** (0.03)	-0.103*** (0.03)
nreprovado	0.137*** (0.00)		0.060** (0.03)		0.058** (0.03)	0.028 (0.03)
apvideo		0.120*** (0.01)		0.045*** (0.00)		
comp_imp		-0.054 (0.04)		-0.016*** (0.00)		
multimid		0.223 (0.15)		0.064*** (0.00)		
sani_den		-0.098 (0.10)		-0.023 (0.02)		
ener_pub				-0.101*** (0.03)		
agua_pub				0.019 (0.02)		
esg_pub		-0.037 (0.12)		-0.015 (0.01)		
lixo_col		0.052*** (0.00)		0.008 (0.01)		
sigma_hat	0.599	0.899	0.226	0.191	0.222	0.223

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

Fonte: Elaboração própria.

Para finalizar a análise da 4ª série, podemos desenvolver uma interpretação contra-factual para o caso de todas as escolas se situarem na fronteira, ou seja, supor que todas são eficientes. Por meio da eficiência Sheppard *output-oriented*, obtemos esse ganho de eficiência multiplicando-se o fator de eficiência pelo nível de insumos utilizado pela observação. Assim, observamos que o ganho das notas de matemática é cerca de 59% (cerca de quatro desvios-padrão), em português cerca de 56% (4,8 desvios-padrão) e no número de alunos, próximo a 58,8%. Quanto ao número de alunos o ganho será factível apenas na existência de muitos alunos fora da escola, não é muito o caso de Minas, contudo, essa análise de projeção tende a desenvolver apenas comparações e interpretações do que nos propõe o modelo.

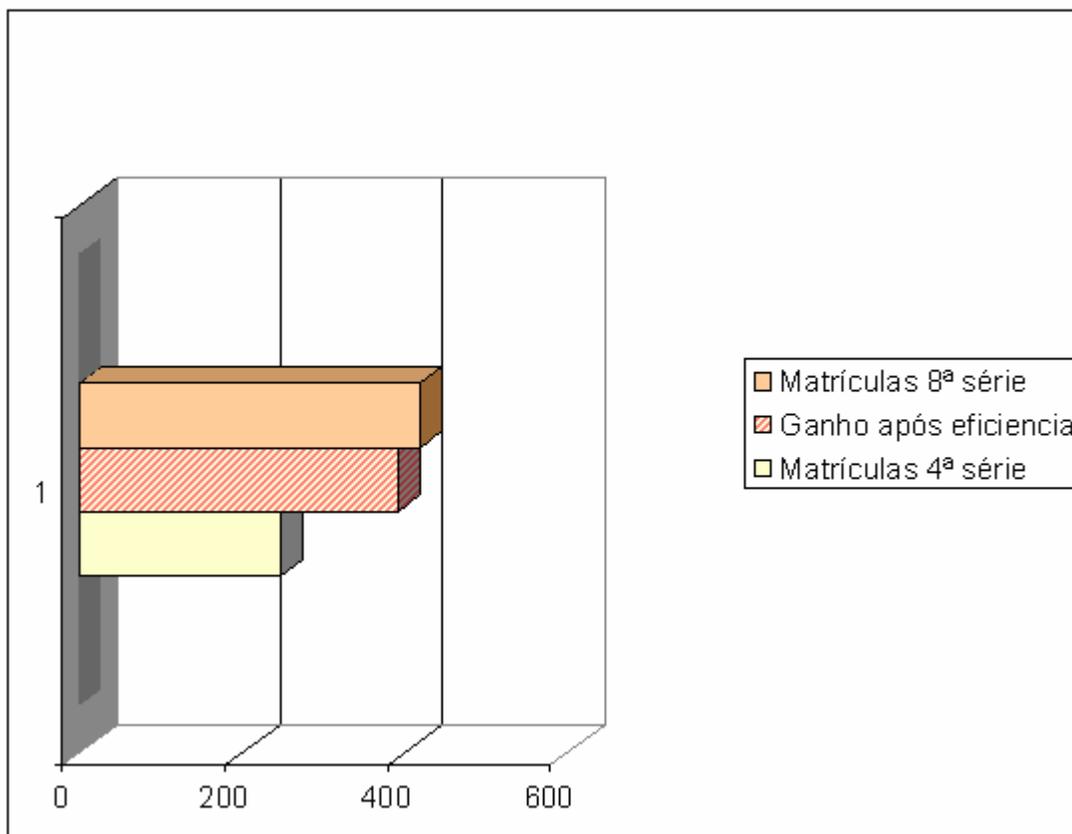
**Gráfico 5.1. Ganho nas Notas pelo critério de Eficiência Output 4ª série.**



Obs.: Nota máxima é o resultado obtido pela proficiência do melhor aluno e não média da melhor turma.

Fonte: Elaboração própria.

**Gráfico 5.2. Ganho na Média de Matrículas pelo critério de Eficiência Output 4ª série.**



Fonte: Elaboração própria.

### 5.1.2. Eficiência 8ª série.

Para a 8ª série temos poucas modificações em relação ao quadro de eficiência apresentado anteriormente. Cabe ressaltar que é a série com maior número de escolas na esfera estadual e de que o ganho nas médias das notas é significativo (mais de 50 pontos, duas vezes o desvio padrão, da escala de habilidades do SIMAVE). A média de eficiência é maior, as escolas se encontram mais perto da fronteira nesta série. 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.

O grupo dos seis primeiros continua o mesmo, sai dele apenas a Região Metropolitana, dando lugar a meso do Sul de Minas. Conseqüentemente, o grupo dos seis últimos ganha apenas um novo membro. Para entender a mudança na posição relativa da Metropolitana é preciso se 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 (segunda posição no *rank3*): são 14 escolas e todas elas na parte decrescente dos ganhos de escala (*decresc*). A correção do *bootstrap* confere um intervalo maior para os eficientes trazendo sua correção para baixo. O modelo Sampaio-Stosic (DEA<sub>ss</sub>) mostra, também, uma queda de posições para esta região mas não tão grande quanto a do *bootstrap*.

Entre os primeiros lugares, a meso de Campo das Vertentes merece novamente destaque, continuou na terceira posição e, na 3ª série, perderá apenas uma 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<sup>53</sup>.

---

<sup>53</sup> O desempenho nessas mesoregiões é acima da média do estado e tal resultado vem seguido por um uso mais comedido dos recursos.

Por isso, apesar de não terem os picos das notas, seus resultados são eficientes. Observamos que no *rank2* essas duas regiões não se desempenham tão bem, isso se deve ao fato de o *ranking* DEA<sub>ss</sub> ser menos suave.<sup>54</sup>

Por outro lado, o desempenho é mais fraco, em termos de eficiência, na mesoregião do Vale do Mucuri, novamente em 11º, assim como o Noroeste de Minas e o Triângulo Mineiro. As duas primeiras apresentam resultados precários em termos de *outputs*. A última possui um nível de insumos elevados, mas resultados fracos, ainda não condizentes com a quantidade de insumos que essa meso possui.

**Tabela 5.10. Ranking das Mesoregiões para a 8ª série.**

Código	Mesoregião	#	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
<b>310013</b>	<b>Minas Gerais</b>	<b>2531</b>	<b>0.755</b>		0.844		<b>0.043</b>		<b>43</b>

Fonte: Elaboração própria.

Os resultados da etapa de dois estágios com os mesmos modelos para as seis regressões são bastante similares para o caso da 8ª série, os sinais dos coeficientes são os mesmos para quase todas as variáveis e iguais para todas as significativas (há uma exceção que será ressaltada). Além disso, há mais variáveis estatisticamente significantes nos modelos que inserem a infra-estrutura escolar (regressões 2 e 4).

<sup>54</sup> De fato, quando olhamos para os insumos mais importantes: custo-aluno, docentes com superior e infra-estrutura, Campo das Vertentes e Oeste de Minas possuem um nível um pouco mais elevado do que o da metropolitana, o que prejudica sua posição no *rank2*, que, apesar de também ser uma classificação corrigida, é um pouco mais sensível do que o *rank1*.

Por serem regressões obtidas por MV, algumas variáveis não atingiam a convergência e não conseguimos uma aproximação dos valores dos coeficientes. Para estes casos, as variáveis que impossibilitavam a convergência estão omitidas (para 8ª série isto ocorre apenas nas regressões 1 e 2).<sup>55</sup>

As variáveis de destaque para a 8ª série são praticamente as mesmas da série anterior, *urbana* se apresenta novamente significativa, no entanto, mesmo quando corrigido para as características de infra-estrutura, o coeficiente continua se apresentando positivo, indicando que escolas urbanas possuem tendência a maior ineficiência. Destacamos, novamente, que, nas escolas rurais, encontramos maior proporção de escolas com rendimentos crescentes, ou seja, uma mesma quantidade de insumos fornece mais resultados quando alocados no ambiente rural.<sup>56</sup>

Duas novas variáveis passaram a ser estatisticamente significantes: merenda escolar (*mere\_esc*), denota que as escolas que possuem merenda escolar são, em média, menos eficientes. Essa variável pode estar indicando uma relativa carência das escolas, são poucas as que dispensam a merenda escolar (menos de 5%), as que o fazem são as que não precisam de merenda, possuem alunos provenientes de famílias mais abastadas. *Cor* se apresenta negativa, indicando que quanto maior proporção de alunos brancos, maior a eficiência, fator explicado também pela disparidade das condições familiares. Essas duas variáveis têm em comum o fato de captarem resultados “residuais”, ou seja, fatores que não estão sendo bem controlados pelas variáveis das condições familiares identificadas pelas variáveis logo abaixo de ‘*mere\_esc*’ e ‘*cor*’. Caso fossem bem controlados, seria de se esperar que estes resultados não fossem estatisticamente significantes, como no caso da 4ª série. Ocorre que a complexidade da análise aumenta de acordo com o que se avança

---

<sup>55</sup> As tabelas descritivas da 8ª série foram suprimidas mas constam no ANEXO E.

<sup>56</sup> Essa explicação é reforçada ao considerarmos as mesmas regressões utilizando a DEA-V na variável dependente ao invés da DEA-C. O coeficiente nas regressões (2) e (4) volta a ser negativo, isso ocorre porque o DEA-V é menos exigente quanto à escala. Quando o DEA-V é empregado, maior número de escolas com rendimentos decrescentes podem ser declaradas eficientes.

nos níveis escolares, indicando que são necessários enfoques sob ângulos mais diversos a fim de controlar essa maior amplitude de fatores.<sup>57</sup>

Destacamos, novamente, a importância da escolaridade da responsável feminina (*esc\_resp\_fem*), de ter uma mini-biblioteca em casa (*livros*), da maior proporção de alunos que lê com frequência (*ler\_livro*), de não trabalhar fora (*ntrab\_fora*) e de realizar o dever de casa (*dever\_mat*). A não-reprovação (*nreprovado*) continua indicando escolas mais fracas e, novamente, significativa apenas na regressão (1).

Na parte de infra-estrutura, a análise de ‘*apvideo*’ permanece. A importância do computador é reforçada para 8ª série, sendo significativa a 1% no modelo mais rigoroso (2) e as demais características seguem as considerações feitas para 4ª série.

---

<sup>57</sup> Essa complexidade se revela em parte pelo número maior de matérias que a escola administra com o passar dos anos. Há, também, que considerar que os alunos se encontram em um estágio mais avançado da curva de aprendizado, portanto, vários outros fatores (como merenda e cor) competem para explicação dos resultados. No modelo (1), mais rigoroso, *mere\_esc* e *cor* não são significativas.

**Tabela 5.11. Resultados das Regressões dois estágios 8ª série.**

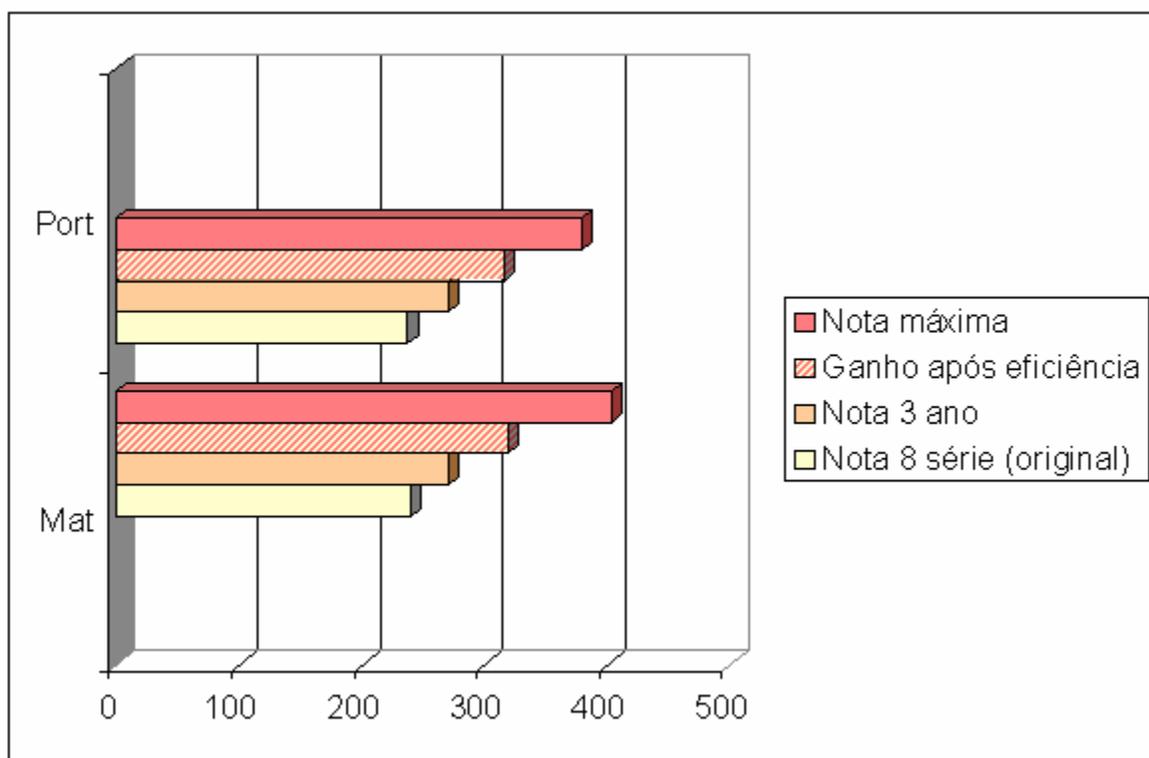
	Regressão (1)	Regressão (2)	Regressão (3)	Regressão (4)	Regressão (5) (tobit1)	Regressão (6) (tobit2)
constante	0.444*** (0.08)	1.202*** (0.01)	1.396*** (0.03)	1.268*** (0.04)	1.397 (0.03)	1.263*** (0.03)
urbana	0.370*** (0.05)	0.019*** (0.00)	0.097*** (0.01)	0.006 (0.01)	0.094*** (0.01)	0.126*** (0.01)
fnde	0.034 (0.11)		0.009 (0.01)		0.009 (0.01)	0.012 (0.01)
mere_esc	0.168 (0.14)		0.044*** (0.01)		0.043*** (0.01)	0.082*** (0.01)
cor	-0.164 (0.21)		-0.076*** (0.02)		-0.074*** (0.02)	-0.045*** (0.02)
esc_resp_fem	-0.414* (0.21)		-0.099*** (0.03)		-0.096*** (0.03)	-0.098*** (0.03)
livros	-0.191 (0.30)		-0.074*** (0.03)		-0.073*** (0.03)	-0.057** (0.03)
internet	0.093 (0.15)		0.025 (0.04)		0.025 (0.03)	-0.005 (0.04)
ler_livro	-0.135 (0.14)		-0.111*** (0.02)		-0.108*** (0.02)	-0.095*** (0.02)
ler_jornais			0.016 (0.02)		0.015 (0.02)	-0.011 (0.02)
ntrab_fora	-0.062 (0.14)		-0.049*** (0.02)		-0.048*** (0.02)	-0.043** (0.02)
ntrab_domestico			-0.012 (0.03)		-0.012 (0.03)	-0.005 (0.03)
devermat	-0.014 (0.16)		-0.044*** (0.02)		-0.043*** (0.02)	-0.054*** (0.02)
nreprovado	0.121*** (0.00)		-0.001 (0.02)		-0.001 (0.02)	-0.012 (0.02)
apvideo		0.011*** (0.00)		0.011*** (0.00)		
comp_imp		-0.008*** (0.00)		-0.008*** (0.00)		
multimid		0.027*** (0.01)		0.027*** (0.00)		
sani_den				0.021* (0.01)		
ener_pub				-0.098*** (0.04)		
agua_pub				0.034*** (0.01)		
esg_pub		0.032*** (0.01)		0.028*** (0.01)		
lixo_col		0.048*** (0.00)		0.042*** (0.01)		
sigma_hat	0.525	0.104	0.112	0.104	0.111	0.123

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

Fonte: Elaboração própria.

O contra-factual da 8ª série está logo abaixo. Ambas as notas, de português e matemática, poderiam aumentar em 33% e o ganho em matrículas é possível ser aumentado em 32%. Ressalva-se que se requer esforço maior na oitava série do que na quarta para incrementar o mesmo número de pontos nas notas por ser já uma série mais avançada. No cenário de todas as escolas sobre a fronteira, a média de notas da 8ª ultrapassaria as da 3ª série. No caso das matrículas, o ganho após a eficiência ultrapassaria o número médio de matrículas da 3ª série inclusive.

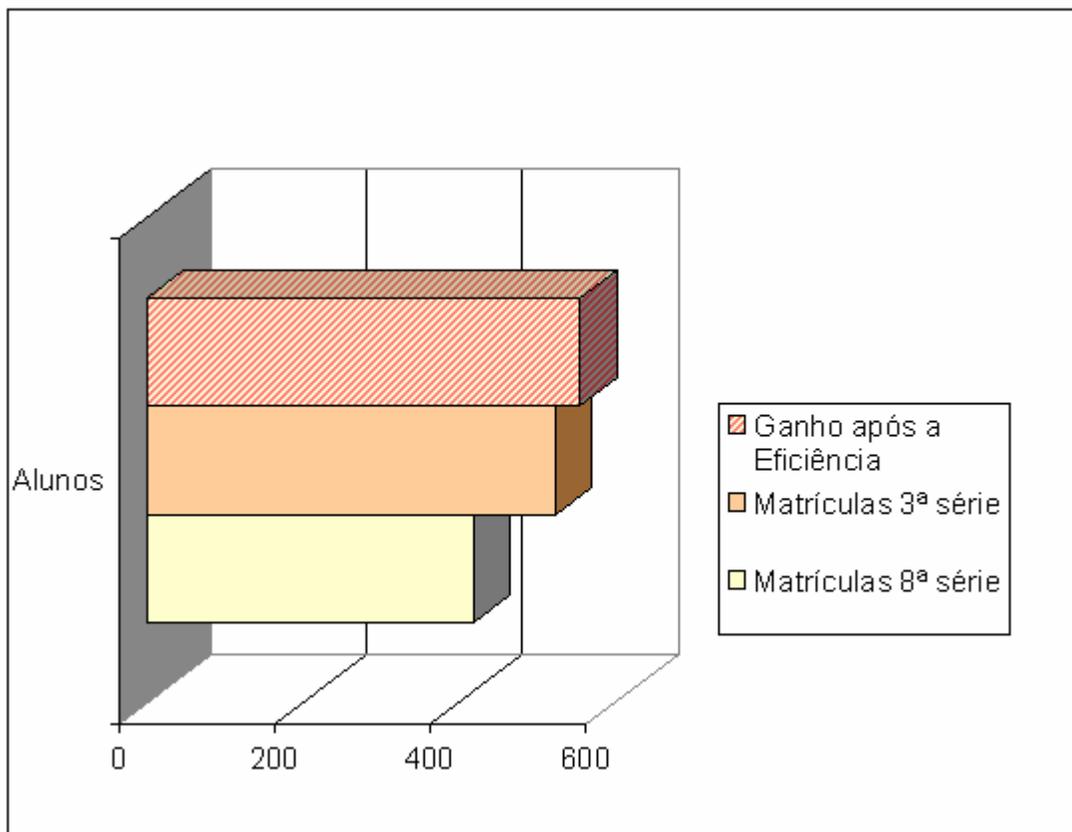
**Gráfico 5.3. Ganho nas Notas pelo critério de Eficiência Output 8ª série.**



Obs.: Nota máxima é o resultado obtido pela proficiência do melhor aluno e não média da melhor turma.

Fonte: Elaboração própria.

**Gráfico 5.4. Ganho na Média de Matrículas pelo critério de Eficiência Output 8ª série.**



Fonte: Elaboração própria.

### 5.1.3. Eficiência 3ª série do ensino médio.

Na terceira série há algumas mudanças de análise 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. Isso indica que a análise ideal passa a ser 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 do que o das séries anteriores. Isso se apresenta em uma média de eficiência DEA-C extremamente baixa (cerca de 0.534), portanto, conjecturamos que se faz necessária uma abordagem mais ampla para a 3ª série. 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 bem menos escolas (1419) e é mais heterogênea que as anteriores.

Um terceiro ponto que distingue a 3ª série é a apresentação de uma elevada aderência entre escolas declaradas eficientes pela SEE/MG e a identificação de eficiência pelo índice Sampaio-Stosic. A correlação de um *ranking* com o outro é de 0.17, o que pode ser considerada elevada já que, das 250 eficientes da lista DEA, mais de 20% (68 escolas), eram, também, consideradas eficientes pela Secretaria, obtidas por outro método.<sup>58</sup>

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.

---

<sup>58</sup> Detalhes sobre a construção da ‘aderência’ no ANEXO D.

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, onde se encontra as de pior desempenho do estado.

**Tabela 5.12. Ranking das Mesoregiões para a 3ª série.<sup>59</sup>**

<b>Código</b>	<b>Mesoregião</b>	<b>#</b>	<b>DEA_v*</b>	<b>Rank1</b>	<b>DEA_ss</b>	<b>Rank2</b>	<b>quocient</b>	<b>Rank3</b>	<b>Decresc</b>
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
<b>310013</b>	<b>Minas Gerais</b>	<b>1419</b>	<b>0.903</b>		<b>0.903</b>		<b>0.052</b>		<b>52</b>

Fonte: Elaboração própria.

No modelo de dois estágios, poucas variáveis das duas primeiras regressões que servem como guia conseguiram atingir a convergência da MV. 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 continua significativa a 1%, indicando que maior escolaridade dos pais colabora com a eficiência dos alunos e da escola.

<sup>59</sup> Novamente as tabelas da 3ª série não estão listadas nessa seção, mas constam no ANEXO E.

**Tabela 5.13. Resultados das Regressões dois estágios 3ª série.**

	Regressão (1)	Regressão (2)	Regressão (3)	Regressão (4)	Regressão (5) (tobit1)	Regressão (6) (tobit2)
constante	0.829*** (0.13)	-2.147 0.01	1.192*** (0.02)	1.109*** (0.06)	1.190*** (0.02)	1.127*** (0.02)
urbana	0.111 (0.12)		0.038** (0.02)	-0.010 (0.02)	0.034** (0.01)	0.073*** (0.02)
fnde			0.000 (0.00)		-0.000 (0.00)	0.001 (0.00)
mere_esc			0.012* (0.01)		0.010* (0.01)	0.021*** (0.01)
cor			-0.076*** (0.01)		-0.071*** (0.01)	-0.059*** (0.01)
esc_resp_fem	-0.531*** (0.00)		-0.071*** (0.02)		-0.064*** (0.02)	-0.088*** (0.02)
livros			-0.024 (0.02)		-0.021 (0.02)	-0.013 (0.02)
internet			0.016 (0.02)		0.015 (0.02)	0.004 (0.02)
ler_livro			-0.053*** (0.01)		-0.049*** (0.01)	-0.047*** (0.01)
ler_jornais			-0.012 (0.01)		-0.011 (0.01)	-0.013 (0.01)
ntrab_fora			-0.024* (0.01)		-0.021* (0.01)	-0.039*** (0.01)
ntrab_domestico			-0.012 (0.02)		-0.011 (0.02)	-0.005 (0.02)
devermat			-0.013 (0.01)		-0.011 (0.01)	-0.022* (0.01)
nreprovado			-0.005 (0.01)		-0.004 (0.01)	-0.008 (0.01)
apvideo				0.004*** (0.00)		
comp_imp		0.012 (0.10)		+0.000 (0.00)		
multimid		0.202 (0.24)		0.007*** (0.00)		
sani_den		0.069*** (0.01)		0.001 (0.01)		
ener_pub				-0.076 (0.06)		
agua_pub				0.023** (0.01)		
esg_pub				0.000 (0.00)		
lixo_col				0.014** (0.01)		
noite			0.029*** (0.01)	0.056*** (0.01)	0.027*** (0.00)	0.023*** (0.00)
sigma_hat	0.233	0.633	0.057	0.057	0.054	0.059

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

Fonte: Elaboração própria.

Para os modelos menos rigorosos como o tobit, *urbana* volta a ser significativa. *Cor* apresenta os mesmos resultados da 8ª série e se confirma novamente a importância da leitura de livros. Os aparelhos de vídeo e TV confirmam as tendências anteriores. Computadores deixam de ser significativos na regressão (2) e (4).<sup>60</sup>

A nova variável acrescentada 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.

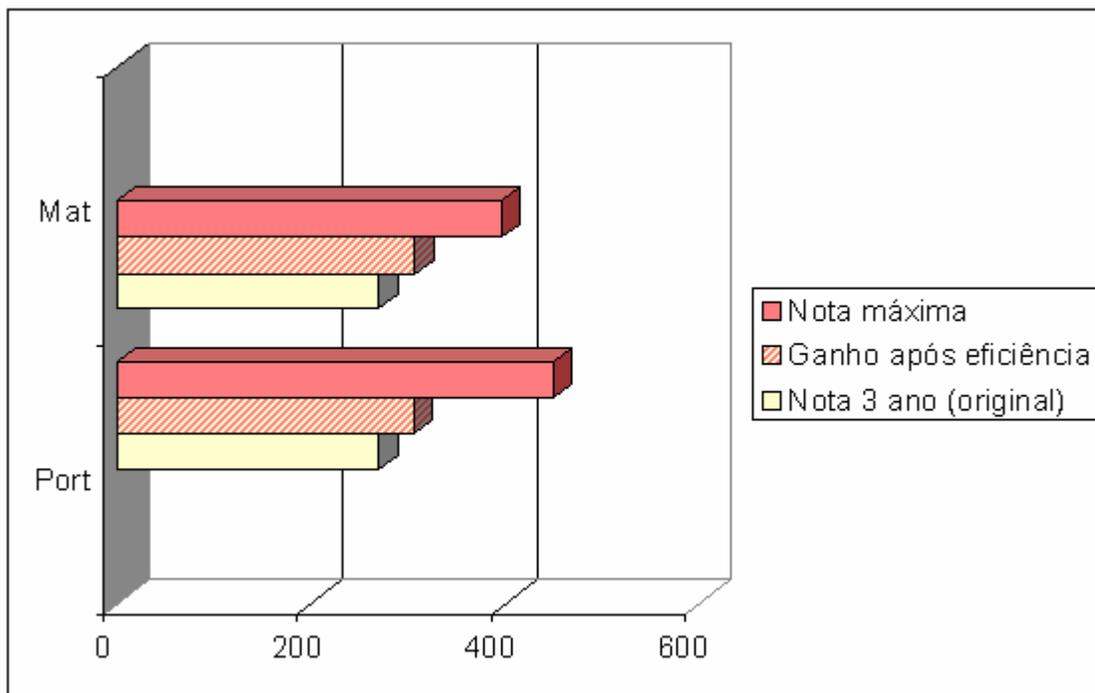
Sobre os ganhos de *output* da 3ª série, o uso da DEA-V parece ser realmente o mais conveniente, porque não exige rendimentos constantes, o que é mais aplicável a alunos na conclusão do ensino médio, onde já houve um acúmulo de conhecimento, estando situados em um ponto de rendimentos decrescentes da curva de aprendizado.

Supondo que todas as escolas são eficientes, os ganhos na prova de matemática e português são por volta de 17% e os ganhos na média de matrículas por turma se situam próximos de 13%. Os valores são menores do que para as demais séries, em virtude da maior incidência de rendimentos decrescentes nessa etapa do ciclo escolar.

---

<sup>60</sup> Quando empregamos novamente o DEA-C ao invés de DEA-V *computador* volta a ser negativo e significante à 1%.

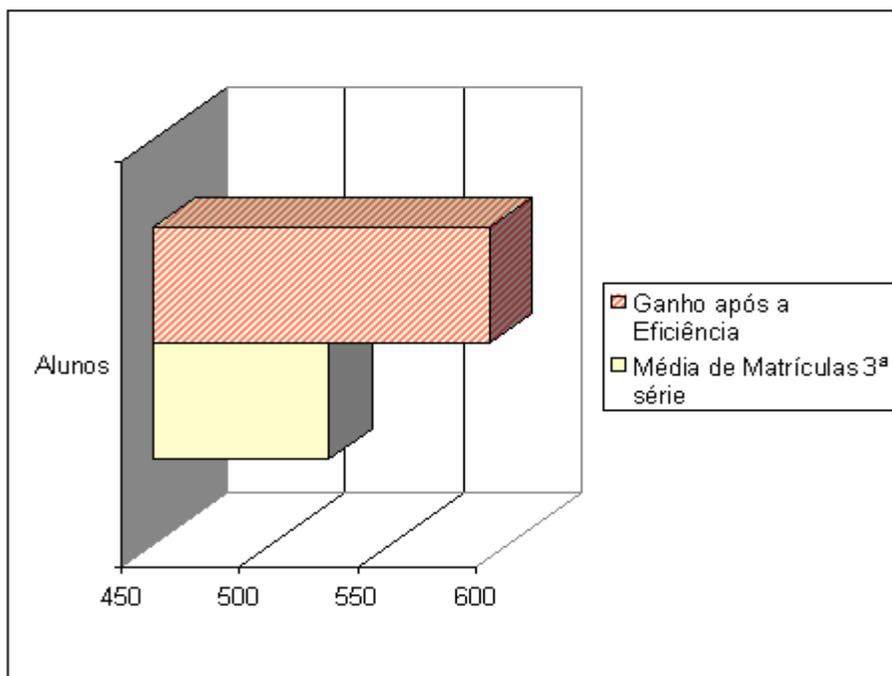
**Gráfico 5.5. Ganho nas Notas pelo critério de Eficiência Output 3ª série.**



Obs.: Nota máxima é o resultado obtido pela proficiência do melhor aluno e não média da melhor turma.

Fonte: Elaboração própria.

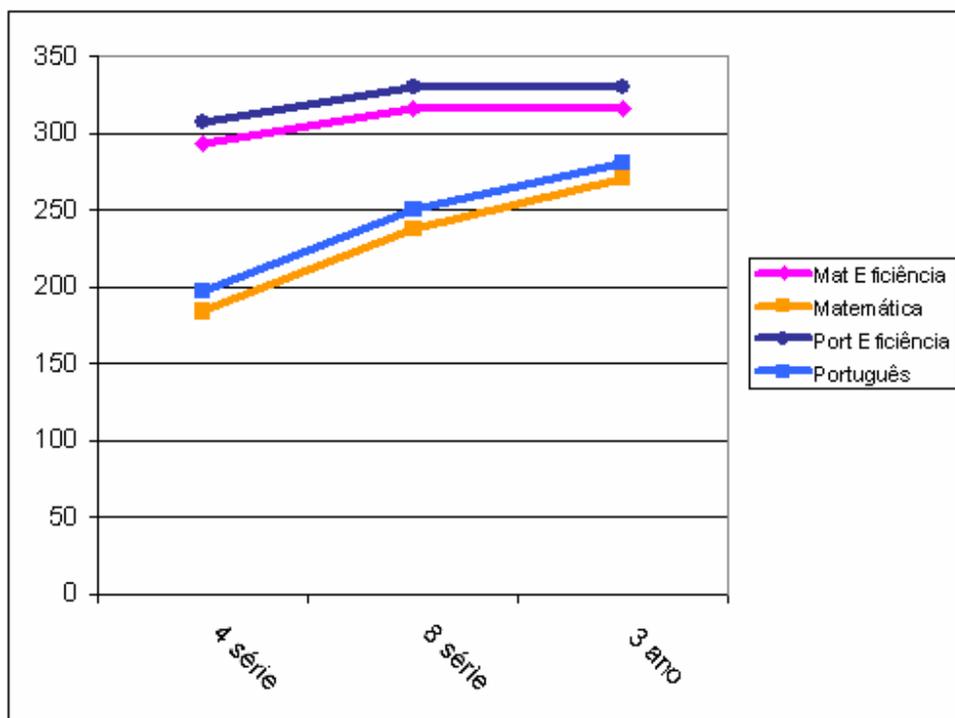
**Gráfico 5.6. Ganho na Média de Matrículas pelo critério de Eficiência Output 3ª série.**



Fonte: Elaboração própria.

Ao final do processo, já com os resultados para as três séries, podemos construir os gráficos da evolução do nível real das variáveis e dos ganhos advindos a partir de uma maior eficiência. Observamos que os ganhos com uma maior eficiência são consideráveis, principalmente para a 4ª série, e que os rendimentos são decrescentes quando temos em conta a fronteira de eficiência. Na fronteira eficiente, da 8ª para a 3ª série praticamente não há ganhos na evolução das notas, esta é uma projeção conservadora, mas está de acordo com a noção de que é mais difícil obter grande progresso nas séries mais adiantadas. A evolução das variáveis realmente observadas não se mantém estável na passagem da 8ª para a 3ª, mesmo assim, seus ganhos são menores do que em relação a mudança de 4ª para 8ª. De todo modo, o que se pode concluir é que entre as retas da fronteira real e as retas do que foi efetivamente realizado existe um espaço onde é possível atuar com melhorias para as escolas. Não é possível que todas escolas sejam eficientes, mas é possível aprimorar as escolas de modo que o desempenho medido dos alunos seja melhor.

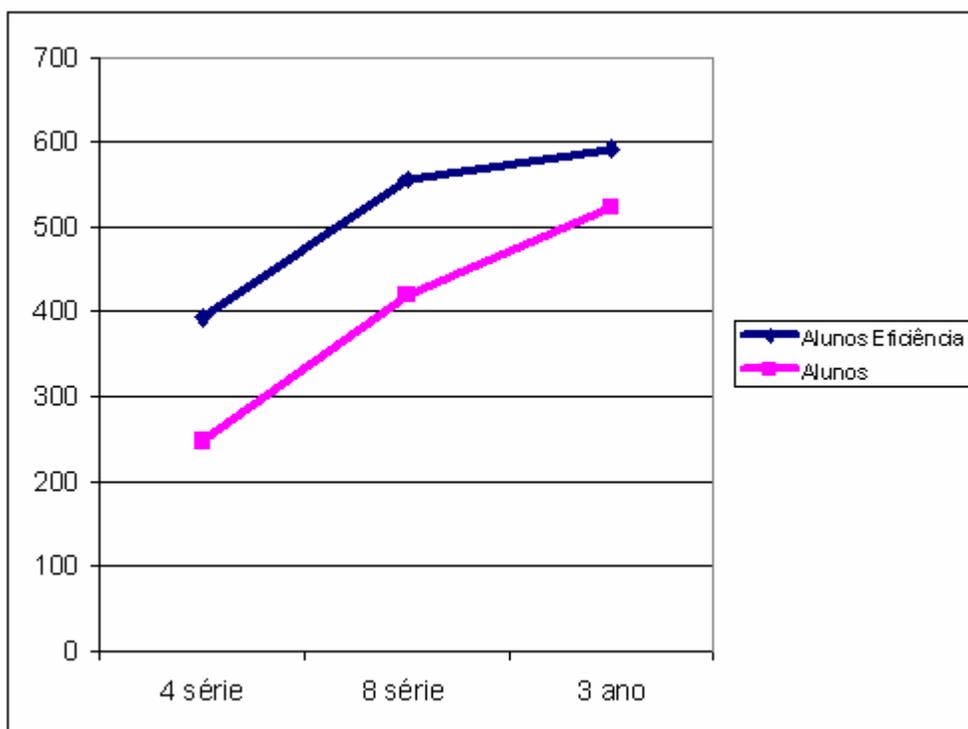
**Gráfico 5.7. Comparativo da Evolução das Notas por Série.**



Fonte: Elaboração própria.

É possível melhorar o ensino aumentando o número de alunos atendidos. Caso haja, no futuro, uma manutenção do crescimento da demanda por educação pública estadual, é possível, tendo em vista o critério de eficiência, aumentar o número de alunos. Isso pode ser evidenciado pelas inúmeras escolas com turmas demasiado pequenas, menos de dez alunos por turma. Assim, em conjunto com uma política de educação geral da população e conscientização de sua importância, conseguir-se-ia aumentar o número de alunos sem reduzir a qualidade do ensino, ao contrário aumentando-a.<sup>61</sup>

**Gráfico 5.8. Comparativo da Evolução das Matrículas por Série.**



Fonte: Elaboração própria.

<sup>61</sup> Há uma questão de cunho prático: é bem provável que a demanda não considere a distribuição de matrículas em cada escola, ou seja, as escolas com poucos alunos se encontram em áreas relativamente mais remotas e a nova demanda pode pressionar justamente as escolas que já estão adequadamente ocupadas. Essa questão sobre a alocação dos alunos é ponto importante a ser visto para o bom uso dos recursos.

## **5.2. Eficiência por Município.**

O âmbito municipal é também importante para compreender o desempenho dos alunos dentro das escolas. Um novo modelo DEA foi construído medindo a eficiência de cada município. Os *outputs* são os mesmos utilizados para a DEA por escolas, porém dados em médias municipais, as matrículas são contabilizadas pelo total de alunos nas escolas estaduais do município. Novos *inputs* foram adotados: população residente, porcentagem da população analfabeta 7 a 14 anos, renda *percapita* R\$ de 2000 e IDH. Dados para o decênio 1991-2000 do Programa das Nações Unidas para o Desenvolvimento (PNUD).<sup>62</sup>

Alguns critérios técnicos foram importantes para o modelo municipal. Os municípios, em conjunto, não rejeitam os retornos constantes de escala, porém, existem, entre eles, dois grupos nítidos. O primeiro e maior grupo é composto de municípios pequenos e médios de baixo desenvolvimento econômico, o segundo grupo é menor, mas formado pelos municípios mais populosos e mais desenvolvidos economicamente, dentre eles a capital, Belo Horizonte.

---

<sup>62</sup> A renda *percapita* é dada para o mês de Agosto do ano 2000, o analfabetismo é dado na proporção inversa, de alfabetizados. Demais informações sobre as variáveis disponíveis no site [www.undp.org.br](http://www.undp.org.br).

Abaixo apresentamos a classificação dos 15 primeiros municípios mineiros de acordo com a DEA. O *ranking* mostra que os municípios mais desenvolvidos confirmam sua polarização também no aspecto educacional. As 15 cidades evidenciadas abaixo concentram cerca de 22% da população do estado e quase 40% da renda além dos maiores índices de desenvolvimento humano.

Na tabela 5.15 acompanhamos que 29% dos 100 primeiros municípios do *ranking* são da Região Metropolitana, outros 56% pertencem às mesoregiões do sul do estado, sendo que as mesoregiões menos desenvolvidas do norte possuem 15%. As escolas eficientes das mesoregiões mais pobres estão representadas no *ranking* dos 100 primeiros pelos municípios de Timóteo (36°) e Pirapora (42°), pertencentes ao Vale do Rio Doce e ao Norte, respectivamente. Ainda assim, são municípios próximos à capital.<sup>63</sup>

**Tabela 5.14. Ranking das cidades com escolas eficientes DEA.**

<b>Cidade</b>	<b>Meso</b>	<b>Escolas*</b>	<b>População</b>	<b>Rendapc</b>	<b>IDH</b>	<b>Custos</b>
Guaxupé	Sul/Sudeste	2	47.036	233.84	0.739	84.32
Nova Lima	Metropolitana	3	64.387	244.91	0.744	86.12
Lagoa Formosa	Triângulo	2	16.293	147.32	0.689	118.65
Belo Horizonte	Metropolitana	98	2.238.526	414.94	0.791	3116.76
Lagoa Santa	Metropolitana	2	37.872	206.49	0.727	84.08
Uberlândia	Triângulo	21	501.214	306.29	0.777	847.46
São João Del Rei	C. das Vertentes	6	78.616	189.03	0.726	265.99
Monte Carmelo	Triângulo	2	43.899	207.24	0.717	79.00
Pedro Leopoldo	Metropolitana	3	53.957	197.24	0.748	123.64
Pará de Minas	Metropolitana	5	73.007	202.17	0.727	160.59
Frutal	Triângulo	3	46.566	233.27	0.731	125.56
Araxá	Triângulo	4	78.997	224.19	0.736	154.33
Itajuba	Sul/Sudoeste	4	84.135	246.67	0.768	159.05
Juiz de Fora	Zona da Mata	27	456.796	311.64	0.768	1000.45
Pouso Alegre	Sul Sudoeste	7	106.776	276.92	0.763	337.34
<b>Total**</b>		189	3.928.077	242.81	0.743	6743.40
<b>Minas Gerais</b>		<b>1435</b>	<b>17.717.220</b>	<b>118.72</b>	<b>0.631</b>	<b>58206.83</b>

\* Escolas do Ensino Médio, referência principal para DEA-Municípios.

\*\* Rendapc e IDH estão dados na média.

Nota: A tabela 5.17 apresenta a eficiência medida por rendimentos constantes, mas tem como referência para obtenção do viés os rendimentos variáveis. São dois níveis de rendimentos ctes, um para cada grupo. Fonte: Elaboração própria.

<sup>63</sup> O *ranking* completo das 100 primeiras, os detalhes da construção dessa classificação e o critério para obtenção dos dois grupos de municípios estão no ANEXO F. Importante ressaltar que esse *ranking* se baseia na 3ª série do ensino médio, já que é a série que abrange o maior número de municípios.

**Tabela 5.15. Distribuição das Mesoregiões nos rankings.**

	<i>Metrop</i>	<i>Norte*</i>	<i>Sul**</i>
100 primeiros do ranking	29%	15%	56%
100 últimos do ranking	10%	36%	54%

\* Norte: Noroeste, Norte de Minas, Jequitinhonha, Mucuri e V. do R. Doce.

\*\* Sul: Triângulo, Central, Oeste, Sul/Sudoeste, C.Vertentes e Zona da Mata.

Fonte: Elaboração própria.

Foram detectados 35 municípios *outliers*. Tais municípios servem de informação já que conseguem resultados muito bons empregando baixos níveis de insumo, no entanto, seu aparecimento distorce a medição para os demais. Abaixo estão relacionados os 14 mais influentes (de maior *leverage*).

**Tabela 5.16. Municípios outliers.**

<b>Cidade</b>	<b>Meso</b>	<b>Escolas*</b>	<b>População</b>	<b>Rendapc</b>	<b>IDH</b>	<b>Custos</b>
São Vicente de Minas	Sul/Sudoeste	1	6.163	149.77	0.681	47.56
Coronel Xavier Chaves	C das Vertentes	1	3.185	89.72	0.663	55.19
Gouveia	Jequitinhonha	1	11.689	99.38	0.662	31.70
Guarani	Zona da Mata	1	8.520	135.30	0.671	64.38
Abaeté	Central	1	22.360	138.06	0.680	46.37
São Jose da Safira	V. do Rio Doce	1	3.894	88.32	0.546	26.05
Carai	Jequitinhonha	1	20.981	57.57	0.526	37.01
São João da Lagoa	Norte	1	4.400	54.52	0.566	27.88
São José do Jacuri	V. Rio Doce	1	6.789	69.58	0.546	57.59
Cedro do Abaeté	Central	1	1.289	99.81	0.649	41.67
São João das Missões	Norte	1	10.230	37.23	0.503	25.37
Itaipé	V. Mucuri	1	10.751	55.66	0.526	28.82
São Sebastião do Rio Preto	Metropolitana	1	1.779	46.61	0.527	55.09
Comercinho	Jequitinhonha	1	10.204	47.76	0.528	50.09
<b>Total**</b>		<b>14</b>	<b>80.879</b>	<b>83.52</b>	<b>0.591</b>	<b>594.77</b>
<b>Minas Gerais</b>		<b>1435</b>	<b>17.717.220</b>	<b>118.72</b>	<b>0.631</b>	<b>58206.83</b>

\* Escolas do Ensino Médio, referência principal para DEA-Municípios.

\*\* Rendapc e IDH estão dados na média.

Fonte: Elaboração própria.

A regressão final destaca os pontos importantes presentes na educação estadual dos municípios mineiros. Seguindo ainda a interpretação na ordem da inversa do índice  $\delta_i$ , ou seja de 0 a  $\infty$ . O resultado de *pmat* está indicando que quanto maior a nota, maior a eficiência. Existe algum pequeno efeito de escala evidenciado pelo pequeno coeficiente significativo de *matriculas\_total*. Os *custos*, apesar da direção correta, não se apresentam significantes. Municípios eficientes tendem ter mais *escolas*, resultado de serem mais desenvolvidos economicamente, por isso também tendem a ter maior renda *percapita* e IDH. As variáveis *Infra*, *população* e *alfabetização* se mostram significantes, mas quanto maior são os níveis de cada uma delas menor a eficiência. Esses resultados, em particular o da alfabetização, indicam que as cidades desenvolvidas ainda concentram grandes níveis de pobreza e podem aproveitar a infra-estrutura que possuem para melhorar os serviços.

**Tabela 5.17. Resultados da regressão DEA-Municípios.**

4ª Série	
constante	1.897*** (0.04)
urbana	-0.142 (0.00)
pmat	-0.3e-03*** (0.00)
pport	-0.002 (0.00)
mat_total	-9.39e-06 (0.00)
custo_total	0.3e-04 (0.00)
total_docente	-3.40e-06 (0.00)
escolas	-.0559*** (0.01)
infra	0.0119*** (0.00)
pop2000	5.54e-07 (0.00)
analf_7_14a	-0.99e-03*** (0.00)
rendapc2000	-0.273e-4 (0.00)
idh	0.1620*** (0.06)
metrop	-0.004 (0.00)

Fonte: Elaboração própria. Regressão Truncada.

O mais importante no terceiro modelo é a divisão dos “modos de produção” entre municípios polarizadores e polarizados. Os grandes municípios se saem bem quando lhes é reservada uma fronteira que considere a sua escala. Vários municípios pequenos se desempenham mal perante os critérios de eficiência, mas isso não impede de obtermos alguns bons exemplos de eficiência como o caso de Lagoa Formosa, do Triângulo Mineiro. Os motivos que levam a essa separação precisam ser investigados mais a fundo, uma hipótese a ser considerada é a de que municípios economicamente mais desenvolvidos possuem maior facilidade no acesso à informação, tornando a obtenção de conhecimento mais rápida. Há, porém, um custo de vida mais elevado e uma maior infra-estrutura que eleva o nível de insumos e transforma a fronteira que se aplica em tais casos. Quando os municípios maiores falham em prover maiores níveis de *output*, sua eficiência despenca.

Assim como os demais serviços, é natural que ocorra a polarização da educação à medida que o ensino se torne mais complexo. O objetivo de uma população mais escolarizada é proporcionar maior capacidade produtiva. A questão importante a ser respondida é se a educação promove de forma igualitária as chances de desenvolvimento da população, e se a educação nos grandes centros promove a capacitação profissional com maior qualidade. Mesmo sendo melhor, a formação educacional dos municípios mais desenvolvidos pode estar aquém da capacidade total já que observamos que os pólos possuem maior desenvoltura para separar o ensino público do privado (o que pode ser verificado pela maior proporção do privado nas cidades maiores)<sup>64</sup>. A consequência dessa separação pode ser um dos motivos para que a formação educacional no Brasil esteja em atraso dado o acesso diferenciado entre ricos e pobres.

---

<sup>64</sup> Anexo F.

## 6. Conclusão.

Nesta dissertação optamos por utilizar o modelo DEA de dois-estágios para desenvolver a análise de eficiência das escolas estaduais do ensino fundamental e médio de Minas Gerais.

As novas bases de dados educacionais do SIMAVE/PROEB e SICA/SEE foram fundamentais para desenvolver a análise de eficiência por escola. O uso da DEA nessa base nos permitiu observar fatores determinantes da eficiência técnica escolar que antes permaneciam encobertos em estudos semelhantes aplicados ao Brasil.

As escolas eficientes fornecem o melhor ensino a seus alunos a um custo significativamente menor do que as outras escolas. Alguns fatores, como a infraestrutura, desempenham papéis importantes para propiciar melhoria dos resultados. Nenhuma escola eficiente *não-outlier* possuía completa ausência de infraestrutura, ou seja, todas as eficientes possuíam ao menos a luz elétrica instalada e alguma forma de esgotamento sanitário, isso quando não possuíam todos os fatores (característica da maioria das escolas eficientes). O computador dentro da escola mostrou-se um bom indicador de eficiência, não que esteja sendo usado diretamente na melhoria dos resultados dos alunos, mas capta alguns fatores indiretos que podem ir desde uma organização ou administração mais eficiente até uma *proxy* de que tais escolas estão situadas em áreas privilegiadas, captando mais recursos e melhores alunos.

A análise de eficiência aponta também materiais subutilizados como televisores e outros aparelhos correlatos (parabólica e vídeo-cassete), além de alguns outros recursos como bibliotecas, videotecas e laboratórios de ciências. Tais insumos não são bem empregados, talvez, por não estarem devidamente equipados para o uso diário.

O recurso mais importante, que não possibilitou captar o efeito na eficiência, foi professores com formação superior. As escolas ineficientes possuem uma proporção maior de professores formados. É claro que a formação superior contribui para melhorar ensino, no entanto, os dados não são suficientes para distinguir de forma relevante as escolas.

Outros indicadores importantes de eficiência são as variáveis familiares do questionário SIMAVE: 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 a leitura e no hábito de estudo, assim como o de fornecer o ambiente mínimo para que os alunos se despenhem 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 permitem facilitar o caminho do aproveitamento dos recursos.

Sendo assim, o papel que cabe às políticas públicas, frente essa análise, é o de focar na complementariedade da formação educacional da população. Para que a escola funcione, é preciso que as demais condições de educação geral, e mesmo recursos materiais das famílias atendam a condições mínimas.

Alcançando, paralelamente, melhores condições para as famílias, e mais recursos para as escolas, a eficiência, e portanto a qualidade do ensino, poderá se desenvolver. Em termos do ganho no produto por meio de uma maior eficiência, vimos que o nível das notas no SIMAVE, e até o número de matrículas podem aumentar se aproveitadas as escolas com capacidade ociosa. É possível dar verdadeiros saltos qualitativos no que os alunos aprendem dentro das escolas e no

que pode ser medido pelo exame do SIMAVE. No caso onde todas são eficientes, a média de notas da 4ª série pode ultrapassar a da 8ª, essa última série pode passar a média do 3º ano do ensino médio, que por sua vez pode aumentar em cerca de 17%.

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, se destaca positivamente o Jequitinhonha e, em algumas análises, o Norte de Minas, por obterem bons resultados mesmo que com poucos recursos. Na questão dos municípios, reforça-se esse padrão das mesoregiões, sendo que os municípios pólo têm tendência à melhor educação e uma fronteira de produção diferenciada.

Por fim, cabe ressaltar que só foi possível estender tanto a análise dos modelos não-paramétricos e o entendimento dos fatores importantes para a eficiência escolar graças a uma série de avanços ocorridos tanto na metodologia DEA, quanto estudos na área da economia da educação aplicados à realidade brasileira. Essa dissertação se apóia nestes novos métodos e pretende afirmar as contribuições para o entendimento do assunto que, felizmente, tanto tem preocupado a sociedade.

## 7. Bibliografia.

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). Disponível em: <<http://www.ecb.int/pub/pdf/scpwps/ecbwp494.pdf>>. Acesso em: 27 fev. 2006.

AIGNER, D. J.; LOVELL, C. A. K.; SCHMIDT, P. Formulation and estimation of stochastic frontier production function models. **Journal of Econometrics**, v. 6, n.1, p. 21-37,1977.

AIGNER, D. J.; CHU, S. F. On estimating the industry production function. **American Economic Review**. v. 58, n. 4, p. 826-839, 1968.

ALBERNAZ A.; FERREIRA, F. H. G.; FRANCO, C. Qualidade e equidade na educação fundamental brasileira **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v. 32, n. 3, p. 453-476, dez. 2002.

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

BARROS, R. P; MACHADO, A. F.; MENDONÇA, R. **Desigualdades da pobreza: estratégias ocupacionais e diferenciais por gênero.** Rio de Janeiro: IPEA, 1997a, 40p. (Texto para discussão; 453)

BARROS, R. P; MENDONÇA, R. **Investimentos em educação e desenvolvimento econômico.** Brasília: IPEA, 1997b. 8 p. (Texto para discussão; 525)

BARROS, R. P; MENDONÇA, R.; SANTOS, D. D.; QUINTAES, G. **Determinantes do desempenho educacional no Brasil.** Rio de Janeiro: IPEA., 2001. 38 p. (Texto para discussão; 834)

BECKER, G. S. **Human capital:** a theoretical and empirical analysis with special reference to education. New York: National Bureau of Economic Research, Columbia University, 1965.187p.

BECKER, G. S.; KEVIN M. MURPHY; ROBERT F. TAMURA. **Human capital, fertility, and economic growth.** Cambridge: National Bureau of Economic Research, 1990, 44p. (Working papers; 3414)

BEHRMAN, J. R.; BIRDSALL, N. The quality of schooling: quantity alone is misleading. **American Economic Review**. v 73, n. 5, p. 928-946, Dec. 1983.

CARD, D.; KRUEGER, A. B. Does school quality matter? returns to education and characteristics of public schools in United States. **Journal of Political Economy**. v. 100, n. 1, p. 1-40, 1992.

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.

COLEMAN, J.; AZRAEL, J.; FRANK, B.; LEON, C.; DEBEUVAIS, M.; FISCHER, J.; CUTTERIDGE, W.; HOSELITZ, B.; KERR, M.; KIRK-GREENE, A.; LANDÈ, C.; LEWIS, J.; MARVICK, D.; OGUNSHEYE, A.; PASSIN, H.; PLATT, W.; SHILS, E.; SUTTON, F. **Education and political development**. Princeton, New Jersey: Princeton University, 1965. 620 p.

DEBREU, G. The coefficient of resource utilization. **Econometrica**, v. 19, n. 3, p. 273-292, July 1951.

EFRON, B.; TIBSHIRANI, R. **An introduction to the bootstrap**. New York: Chapman and Hall, 1993. 436 p.

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.

FARE, R.; GROSSKOPF, S.; WEBER, W. Measuring school district performance. **Public Finance Quarterly**, v. 17, n. 4, p.409-420, Oct. 1989.

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.) Disponível em: [http://getinternet.ipea.gov.br/ipeacaixa/premio2006/docs/trabpremiados/IpeaCaixa2006\\_Profissional\\_02lugar\\_tema01.pdf](http://getinternet.ipea.gov.br/ipeacaixa/premio2006/docs/trabpremiados/IpeaCaixa2006_Profissional_02lugar_tema01.pdf). Acesso em: 02 jan. 2007.

FELÍCIO, F.; FERNANDES, R. O efeito da qualidade da escola sobre o desempenho escolar: uma avaliação do ensino fundamental no Estado de São Paulo. In: ENCONTRO

NACIONAL DE ECONOMIA, 33., 2005, Natal. **Anais**. Belo Horizonte: ANPEC, 2005. (Disponível em CD-ROM)

FERRIER, G. D. Ownership type, property rights, and relative efficiency. In: CHARNES, A.; COOPER, W.; LEWIN, A.Y.; SEIFORD, L. M. (Eds.), **Data envelopment analysis: theory, methodology and applications**. Boston: Kluwer Academic, 1994. p. 273-283.

FLETCHER, P. R.; RIBEIRO, S. C. **Modeling education system performance with demographic data: an introduction to the PROFLUXO model**. Brasília: IPEA, 1989. 20 p. (Mimeogr.).

GARCIA, F.; PONS, T.; MUSSOLINI, C. Os efeitos da qualidade do ensino sobre o crescimento econômico. In: ENCONTRO NACIONAL DE ECONOMIA, 33., 2005, Natal. **Anais**. Belo Horizonte: ANPEC, 2005. (Disponível em CD-ROM)

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.

GOLGHER, A. Modelo profluxo e indicadores derivados. In: RIANI, J; RIOS-NETO, E. (Org.). **Modelo profluxo e indicadores derivados**. Campinas: ABEP, 2004, p.159-208.

GOMES, C. A.; SOBRINHO, J. A. (Org.). **Qualidade, eficiência e equidade na educação básica**. Brasília: IPEA, 1992. (Trabalho apresentado no seminário Qualidade, Eficiência e Equidade na Educação Básica, nov. 1991)

GREENE, W. H. **Simulated likelihood estimation of the normal-gamma stochastic frontier function**. New York: New York University, Stern School of Business, 2000. 18 p. (Working papers; 00-05). Disponível em:  
<<http://www.stern.nyu.edu/eco/wkpapers/workingpapers00/00-05Greene.pdf>>

GREENE W. The econometric approach to efficiency analysis. In: FRIED, H.; LOVELL, C. A. K.; SCHIMDT, P. (Eds). **The measurement of productive efficiency: techniques and applications**. Oxford: Oxford University, 1993. p.68-119.

HANUSHEK, E. A.; LUQUE, J. **Efficiency and equity in schools around the world**. Cambridge, Massachusetts: National Bureau of Economic Research, 2002, 42 p. (Working Papers; 8949).

HANUSHEK, E. A., The trade-off between child quantity and quality. **Journal of Political Economy**. Chicago, v. 100, n. 1, p. 84-117, Feb.1992.

HANUSHEK, E. A. The economics of schooling: production and efficiency in public schools **Journal of Economic Literature**, v. 24, n. 3, p. 1141-1177, Sept. 1986.

KNEIP, A.; SIMAR, L.; WILSON, P. W. **Asymptotics for DEA estimates in nonparametric frontier models**. 2003. 37 p. (Technical report; 0323). Disponível em:  
<<http://www.stat.ucl.ac.be/ISpub/tr/2003/TR0323.pdf>>. Acesso em: 26 abr. 2006.

KRUEGER, A. B.; LINDAHL, M.. Education for growth: why and for whom? **Journal of Economic Literature**, v. 39, n. 4, p. 1101-1136, 2001.

LAU, L. D.; JAMISON, F. LOUAT. **Education and productivity in developing countries: an aggregate production function approach**. Washington: World Bank, 1991, 38 p. (Working papers; 612)

MACHADO, A. F.; MORO, S.; MARTINS, L.; RIOS, J. Qualidade do ensino em matemática: determinantes do desempenho de alunos em escolas públicas estaduais mineiras. In: ENCONTRO NACIONAL DE ECONOMIA, 34., 2006, Salvador. **Anais**. Belo Horizonte, ANPEC, 2006. (Disponível em CD-ROM)

MCCARTY, T A.; YAISAWARNG, S. Technical efficiency in New Jersey School Districts. In: FRIED, H. O; LOVELL, C. A. K.; SCHMIDT, S. S. (Eds). **The measurement of productive efficiency: techniques and applications**. Oxford: Oxford University, 1993. p.271-287.

MINCER, J. On-the-job training: costs, returns, and some implications. **Journal of Political Economy**, v. 70, n.5 Suppl., p. 50-79, 1962.

MENEZES-FILHO, N. A. **Qualidade das escolas, dos alunos ou dos professores?: diferenças nos resultados do SAEB entre 1995 e 2001, entre escolas públicas e privadas e entre regiões do Brasil**. São Paulo: Universidade de São Paulo, 2004. (Mimeogr.).

BRASIL. Ministério da Educação e Cultura. Instituto. Nacional de Estudos e Pesquisas Educacionais. **Censo educacional 2003: avaliação**. Brasília: INEP, 2003.

MORAN, J. M; MASSETO, M. BEHRENS, M. **A utilização inovadora da televisão, Internet e outras técnicas**. Campinas: Papyrus, 2002.

MORAN, J. M. Desafios da televisão e do vídeo à escola. (Texto de apoio ao programa Salto para o Futuro da TV Escola no módulo TV na Escola e os Desafios de Hoje) Disponível em: <<http://www.eca.usp.br/prof/moran/desafio.htm>>. Acesso em: 31 fev. 2007.

OLIVEIRA, J.B. A.; SCHWARTZMAN, S. **A escola vista por dentro**. Belo Horizonte: Alfa Educativa, 2002.

PEREIRA, D.R.M. **Fatores associados ao desempenho escolar nas disciplinas de matemática e de português no ensino fundamental: uma perspectiva longitudinal**. 2006. Tese (Doutorado em Demografia). Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

PRITCHETT, LANT.. **Where has all the education gone?** Washington: World Bank, 1996, 56 p. (Working papers; 1581). Disponível em: <<http://www-wds.worldbank.org>>. Acesso em: 03 fev. 2006.

RIANI, J. **Determinantes do resultado educacional no Brasil: famílias, perfil escolar dos municípios e dividendo demográfico numa abordagem hierárquica e espacial**. 2005. 234 f.

Tese (Doutorado em Demografia). Centro de Desenvolvimento e Planejamento Regional, Universidade Federal de Minas Gerais. Disponível em:  
<[http://www.cedeplar.ufmg.br/demografia/teses/2005/Juliana\\_Riani.pdf](http://www.cedeplar.ufmg.br/demografia/teses/2005/Juliana_Riani.pdf)>. Acesso em: 10 nov. 2006.

RIANI, J. L.; RIOS-NETO, E. L. G.; (Org.). **Introdução à demografia da educação**. Campinas: ABEP, 2004. 212 p.

RIOS-NETO, E. L. G.; RIANI, J. L. R.; CÉSAR, C. C. **Introdução à demografia da educação**. Campinas: ABEP, 2004. Parte 4, Mother's or teacher's education?: educational stratification and grade progression in Brazil. (Disponível em CD-ROM)

SALA-I-MARTIN, X. **15 years of new growth economics**: what have we learnt. Santiago del Chile: Central Bank of Chile, 2002. 27 p. (Working papers; 172) Disponível em:  
<<http://www.bcentral.cl/eng/stdpub/studies/workingpaper/pdf/dtbc172.pdf>>. Acesso em: 10 jun. 2006.

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.

SEIFORD, L. M.; THRALL, R. Recent developments in DEA, the mathematical programming approach to frontier analysis. **Journal of Econometrics**, v. 46., n.1-2, p. 7-38, 1990.

SENA, V. The frontier approach to the measurement of productivity and technical efficiency. **Economic Issues**, v. 8, n. 2, p.71-97, Sept. 2003.

SHEPPARD, R. W. **Theory of cost and production function**. Princeton, NJ: Princeton University, 1970. 308p.

SICA: Sistema Informacional Custo Aluno da SEE/MG. **Site**. Disponível em:  
<<http://www.fundep.ufmg.br/homepage/cases/517.asp>> e  
<[http://www.prodemge.gov.br/solucoes\\_tecnologicas\\_executivo.asp](http://www.prodemge.gov.br/solucoes_tecnologicas_executivo.asp)>. Acesso em: 16 jan. 2007.

SILVERMAN, B. W. **Density estimation for statistics and data analysis**. New York: Chapman & Hall, 1986. 175 p.

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**. Louvain-la-Neuve, Bélgica: Institut de Statistique,

Université Catholique de Louvain. 2003. (Discussion paper) Disponível em: <<http://www.stat.ucl.ac.be>>. Acesso em: 28 abr. 2006.

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. Statistical inference in nonparametric frontier models: the state of the art. **Journal of Productivity Analysis**, v. 13, n. 1, p. 49-78, 2000.

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.

SIMAVE: Sistema Mineiro de Avaliação do Ensino Público. **Site**. Disponível em: <<http://www.simave.ufjf.br/home.asp>>. Acesso em: 28 abr. 2006.

SOARES, T. Utilização da teoria da resposta ao item na produção de indicadores sócio-econômicos. **Pesquisa Operacional**, v. 25, n. 1, p. 83-112, jan./abr. 2005.

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. 2005a. 40 p. Disponível em: <[http://ueg.eerc.kiev.ua/conf/papers/Wilson\\_Paul.pdf](http://ueg.eerc.kiev.ua/conf/papers/Wilson_Paul.pdf)> Acesso em: 23 jan. 2007

WILSON, P. W. FEAR 1.0 a software package for frontier efficiency analysis with r. Austin, Texas: Department of Economics, University of Texas at Austin, 2005b, 13 p. Disponível em: <<http://www.clemson.edu/economics/faculty/wilson/courses/bcn/papers/fear.pdf>>.

WILSON, P. W. Detecting influential observations in data envelopment analysis. **Journal of Productivity Analysis**, v. 6, n. 1, p. 27-45, Apr. 1995.

ZELENUYK, V. **Power of signifiante test of dummies in Simar-Wilson two-stage efficiency analysis model**. Louvain-la-Neuve, Bélgica: Institut de Statistique, Université Catholic de Louvain. 2003, 5 p. (Thecnical report; 0539). Disponível em: <<http://www.stat.ucl.ac.be/ISpub/tr/2005/TR0539.pdf>>. Acesso em: 05 abr. 2006.

## Anexo A (Métodos)

### A.1. Método de Fronteiras Estocásticas.

O método de fronteiras estocásticas é uma abordagem alternativa para a construção de fronteiras de eficiência. Os índices que dele derivam também possuem base no trabalho de Farrel (1957). No entanto, a fronteira não é mais construída por programação linear e sim através de meios econométricos.

Aigner e Chu (1968) identificaram a importância de uma correção para a tradicional análise de estimação de fronteiras. O procedimento utilizado pelos dois autores é econométrico, porém, fazem uso de uma correção por programação linear de forma a incorporar apenas os erros positivos. A equação de Aigner e Chu:

$$\ln(y_i) = x_i \beta - u_i \quad (16)$$

Os  $\beta$ 's são obtidos por mínimos quadrados, no entanto é aplicada a restrição tal que  $\hat{A}x_1^{\hat{\alpha}} x_2^{\hat{\beta}} \geq x_0$ , essa restrição faz com que os erros ( $u_i$ ) sejam sempre não-negativos, tendo-se o resultado de uma fronteira de eficiência máxima. A Farrel-eficiência ( $\theta_i$ ) é obtida por meio de  $\exp(-u_i)$ .

A lógica dos erros aleatórios estocásticos não-negativos segue-se em trabalhos posteriores. Aigner Lovell e Schmidt (1977) o separaram, especificando por ( $u_i$ ), a variável randômica não negativa e de ( $v_i$ ) a parte dos erros normais. Como já antecipado na equação (3) do capítulo anterior:

$$\ln(y_i) = x_i \beta + v_i - u_i \quad (17)$$

A distribuição dos erros não negativos ( $u_i$ ) é geralmente tomada como uma exponencial ou uma normal partida (STEVENSON, 1980 apud: COELLI et ali,

1998). Greene (1990) adota uma distribuição Gama de dois parâmetros para esses erros, ganhando alguma generalidade para a distribuição dos erros.

O método mais utilizado para a estimação de (17) é a máxima verossimilhança (GREENE, 1990, 2000; COELLI et alii, 1998), de fácil implementação a partir dos novos avanços computacionais e já disponíveis em diversos softwares estatísticos.

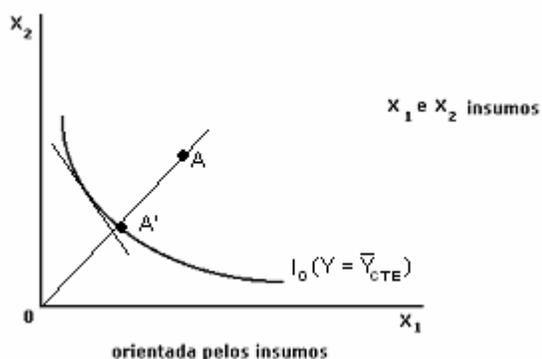
O método de fronteiras estocásticas dispensa o procedimento de dois-estágios. Desde que não correlacionadas com os insumos, as variáveis de *background* familiar podem ser usadas diretamente na equação (17).

Outra vantagem dos procedimentos estocásticos é a de que estão previstas a ocorrência de erros, e tais erros podem ser tratados. Além disso, a fronteira não fica tão sensível ao aparecimento de *outliers*. É possível contar com o cálculo de elasticidade dos parâmetros de acordo com os insumos, o que não pode ser realizado nos métodos não-paramétricos. As desvantagens são referentes à forma funcional. Ela precisa ser especificada anteriormente e não há como, a não ser pelo cálculo de um índice que os agregue, calcular a eficiência para mais de um produto ao mesmo tempo.

## A.2. Construção dos índices e Comparação Sheppard x Farrel.

Para ilustrar a obtenção da eficiência *input-oriented*, exemplificamos por meio de observações com dois insumos e um produto em uma **isoquanta** de eficiência no  $\mathcal{R}^2$ .<sup>65</sup>

**Gráfico A.1. Medição da Eficiência orientação do insumo.**



A isoquanta  $I_0$  indica um nível de produto constante qualquer,  $Y_{CTE}$ , para o qual se quer minimizar custos. O ponto  $A$  é a observação real e  $A'$ , a observação projetada. O índice Farrel-eficiência é obtido por:

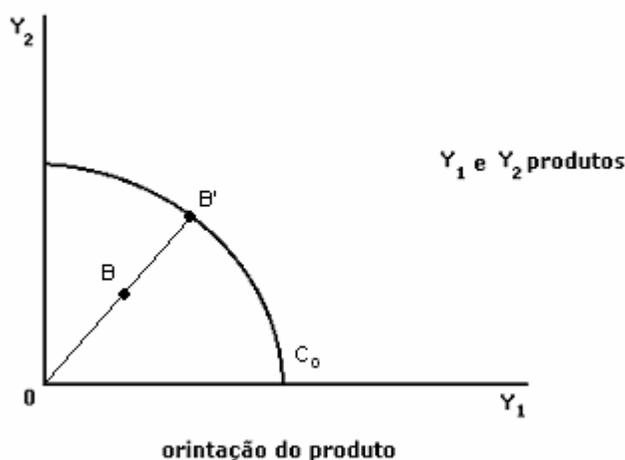
$$\theta_i = \frac{OA'}{OA} \quad (18)$$

Esse índice variará de 0 a 1. Percebemos que 1 é o caso onde  $OA' = OA$  e a observação é ela mesma eficiente. A equação (9) é *input-oriented* porque para se atingir a isoquanta é preciso reduzir radialmente os insumos de maneira que  $x'_1 = \theta x_1$  e  $x'_2 = \theta x_2$ .

<sup>65</sup> É possível obter a eficiência de preços, obtida pela reta da relação de preços representada no gráfico (3.4), chamada de eficiência alocacional, mas não entraremos em detalhes aqui pois essa eficiência não será obtida.

A construção output-oriented pode ser melhor exemplificada no  $\mathfrak{R}^2$  por meio da **curva de possibilidades de produção** onde, dada uma carga de insumos fixa, é preciso expandir um ou os dois produtos para atingir a eficiência:

**Gráfico A.2. Medição da Eficiência orientação do insumo.**



O ponto B é a observação e B' sua projeção,  $C_0$  é a curva de possibilidades de produção máxima. O índice **Farrel-eficiência** será dado por:

$$\theta_i = \frac{OB'}{OB} \quad (19)$$

Observe que o Farrel-eficiência na orientação do produto varia de 1 a  $\infty$ . Existe um outro índice proposto por Sheppard (1970) que consiste no inverso do índice de Farrel. Nesse trabalho, usamos por conveniência o **Sheppard-eficiência** na orientação do produto, que varia de 0 a 1. Para diferenciar do índice de Farrel usa-se o  $\delta_i$  para a Sheppard-eficiência.

Abaixo a fórmula e o quadro resumem a diferença entre os dois índices:

$$\theta_i = \frac{1}{\delta_i} \quad (20)$$

	<b>Orientação</b>	
	<b>Insumo</b>	<b>Produto</b>
<b>Farrel</b> $\theta_i$	0 a 1	1 a $\infty$
<b>Sheppard</b> $\delta_i$	1 a $\infty$	0 a 1

Para fins de programação de softwares, a informação que damos como entrada para a construção do índice é a de que ele deve partir da origem e ajustar sempre a distância à observação real no denominador, em contrapartida, a projeção é posta sempre no numerador, o que constrói o índice Farrel. Para o índice Sheppard deve se proceder ao contrário: a distância à observação real entra sempre no numerador, enquanto que a distancia à projeção pertence ao denominador do índice. Alguns autores não fazem essa distinção para os índices (Coelli et Ali. 1998, p. 158-9), sendo que então, o *Farrel-eficiência* permanece sempre delimitado de 0 a 1, mesmo quando se refere à orientação do produto. No entanto, para a construção do índice em coordenadas polares, a distinção é importante.

## **Anexo B**

Como ressaltado no capítulo 4, alguns poucos valores para as notas de português tiveram de ser imputados. Esse procedimento é realizado pelo comando *impute* do STATA 9.0. O comando apenas completa os *missings*, as demais informações, já presentes da prova de português, são mantidas.

De acordo com o *Help* do programa:

[**impute** organizes the cases by patterns of missing data so that the missing-value regressions can be conducted efficiently; this necessitates a limit of 31 variables in indepvars.]

A imputação é realizada segundo padrões dos dados faltantes, de modo a guiar a regressão que imputará os valores *missings*, obtendo assim a regressão mais eficiente, ou seja, a de mais rápida convergência para a variância mínima dos parâmetros.

Cabe lembrar que a variável imputada *pport* foi utilizada no primeiro estágio, do modelo não paramétrico da DEA e não na regressão. Para o caso não paramétrico de fronteiras, a imputação pode inserir também distorção nos dados, no entanto, isso somente para o caso de várias observações influentes que destoam da média, não é o caso da prova de português, que tem um lastro bastante confiável, que é a prova de matemática.

Mesmo sendo poucas escolas com *missing* em *pport*, julgamos ser mais salutar mantê-las pela imputação do que excluí-las, já que na análise de eficiência é sempre bom identificar escolas específicas, e quanto maior a base de comparação, melhor.

Para 4ª série, então, foram 140 escolas (5.7% da amostra) controladas pelas variáveis:

```
impute pport urbana logmat_1a4 logmat_total cdentro cdentrof logc logcf qdocente
fnde liv_ser4 liv_ser8 mere_esc apvideo comp_imp multimid salas sani_den ener_pub
agua_pub esg_pub lixo_col vcm salee, generate(depvar) nomissings(pmate)
```

5.70% (140) observations imputed

Para 8ª série, 86 escolas (3,4% da amostra):

```
impute pport urbana logmat_5a8 logmat_total cdentro cdentrof logc logcf
qdocente fnde liv_ser4 liv_ser8 mere_esc apvideo comp_imp multimid salas
sani_den ener_pub agua_pub esg_pub lixo_col vcm salee, generate (pporte)
nomissings(pmate)
```

3.40% (86) observations imputed

3ª série, 79 escolas (5,57% da amostra):

```
impute pport codesc urbana logmat_1a4 logmat_total cdentro cdentrof
cdentrom logc logcf logcm qdocente fnde liv_ser4 liv_ser8 mere_esc
apvideo comp_imp multimid salas sani_den ener_pub agua_pub esg_pub
lixo_col vcm salee, generate(pteste) nomissings (pmate)
```

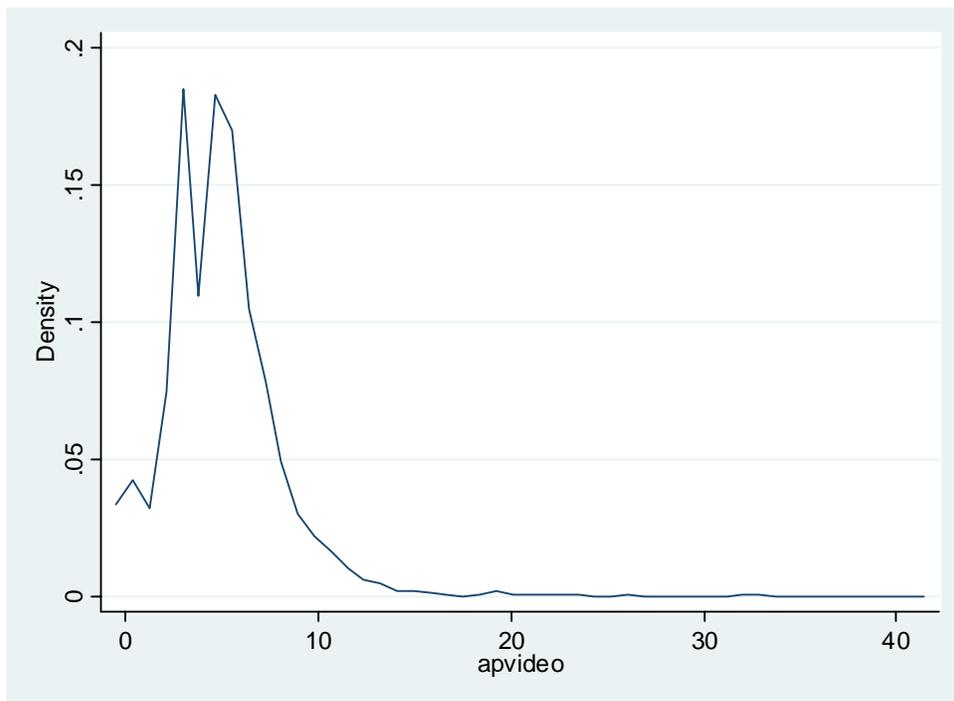
5.57% (79) observations imputed

## Anexo C

Antes de criar o índice VCM, analisamos as três variáveis que o compõem. A primeira delas é *apvídeo*, capta o número de Aparelhos relacionados à Televisão obtida através da soma das variáveis do censo escolar: *vvideo* (número de vídeos-cassete), *vtv* (número de televisores) e *parabol* (número de antenas parabólicas). Havia algumas poucas inconsistências, ou, no mínimo, aparelhos sem uso, já que reportam-se casos onde a escola possui Parabólica mas não tem tv, ou então possui videocassete mas não televisão. Nesse último caso os respondentes aos questionários podem ter tomado vídeo por tv. Em todo caso, acreditamos ser uma variável segura já que são poucos casos onde estas situações reportadas ocorrem.

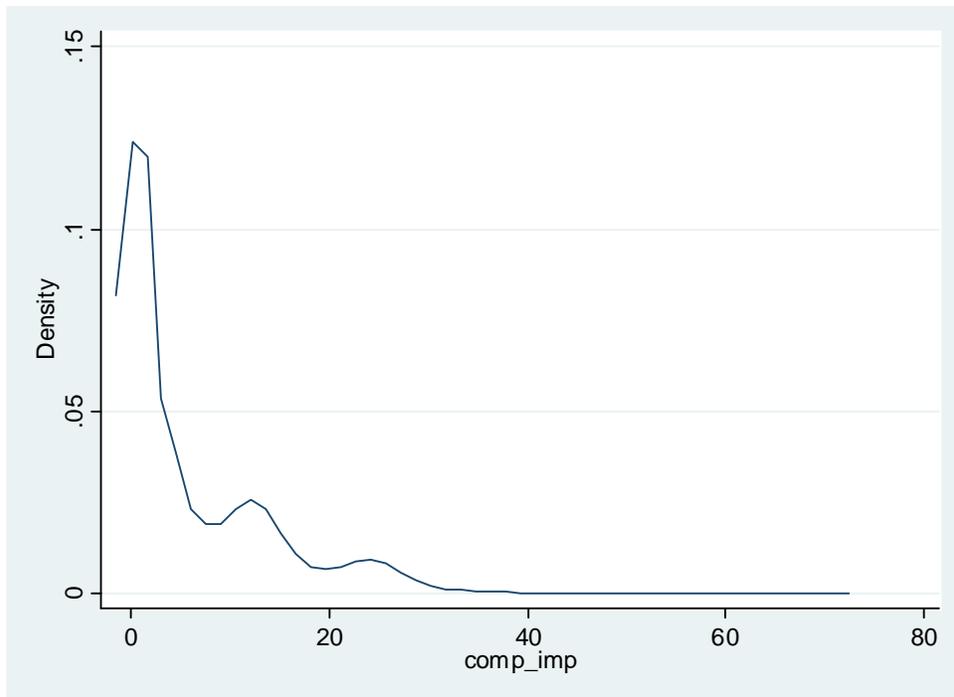
A densidade de Kernel para *apvídeo*, 4ª série como exemplo, está retratada no gráfico abaixo:

**Gráfico C.1. Densidade de Kernel para ‘apvídeo’ (4ª série).**



A Segunda variável, *comp\_imp*, verifica o número de aparelhos de informática que a escola possui. São agregados computadores (qualquer tipo) e impressoras. As variáveis originais do censo somadas são: "*compent*", "*pc486386*", "*compoutr*" e "*vimpress*". No caso dos equipamentos de informática, não ocorre inconsistência ou não-utilização por não ter o complementar, em outras palavras, todas as escolas que reportaram uma impressora disseram ter ao menos um computador, não ocorrendo o caso de impressora sem computador. Outra observação é de que algumas escolas possuíam número elevado de pc's (20 por exemplo), mas nenhuma impressora, o que é improvável, é possível que o respondente, nesses casos, esteja incluindo os pc's como recursos disponíveis aos alunos apenas. Já nas escolas menores, deve haver a diferença de incluir computadores que, na verdade, estão disponíveis apenas à administração e não aos alunos. De qualquer maneira, incorporamos a variável de forma a captar outras componentes por trás, que não só a presença de material de informática.<sup>66</sup>

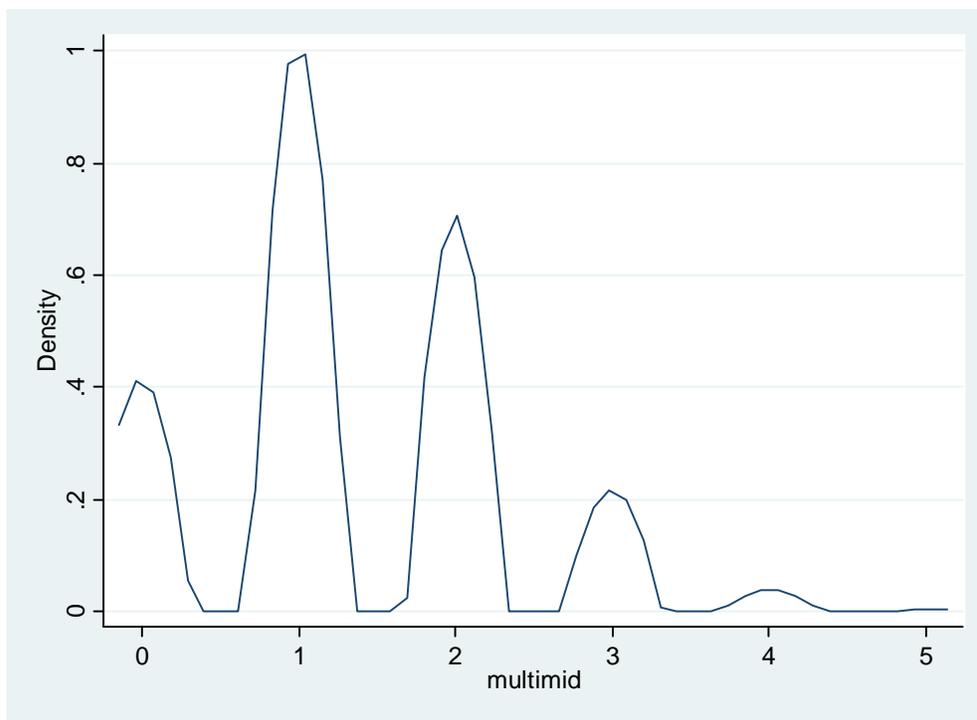
**Gráfico C.2. Densidade de Kernel para 'comp\_imp' (4ª série).**



<sup>66</sup> Mas se desejarmos captar o acesso aos alunos somente, podemos adotar um critério arbitrário onde valha essa variável acima de um valor acima de 4 ou 5, por exemplo, casos onde os pc's devem estar, realmente, sendo aplicados aos alunos. É, de todo modo, arbitrário!

A terceira e última componente de VCM é *mutimídia*, que foi criada para captar os recursos dentro da escola disponíveis aos alunos. Abarca a soma das variáveis originais: "*biblioteca*", *videotec* (videoteca), *saltvvid* (Sala de Tv e Vídeo), *sala\_lei* (Sala de Leitura) e *lab\_info* (laboratório de informática). Observamos pelo gráfico B.3 que essa última variável é discreta, assumindo valor 5 quando a escola possui todos os recursos (menos de 1% para quarta série).

**Gráfico C.3. Densidade de Kernel para 'multimid' (4ª série).**



Após o estudo das três principais variáveis agregadas abordadas acima, concluímos pela importância da decomposição por componentes principais. A distribuição de cada uma delas é bastante assimétrica com longas caudas à direita e também multimodal, além do fato de em algumas distribuições haver muitas ocorrências perto do limite inferior, ou perto do superior, o que, com certeza, influenciaria muito no índice de eficiência, principalmente no caso do inferior (por ser um insumo). Como as variáveis estão correlacionadas é possível construir um índice que as agregue em uma só.

O STATA apresenta um comando para a obtenção do *Principal Component Analysis* – Análise de Componentes Principais, o *pca*. O método de componentes principais objetiva ortogonalizar as informações de forma que as mesmas se situem à menor distância ortogonal dos eixos transformados. Em outras palavras, o método translada os eixos de maneira que eles fiquem mais próximos ortogonalmente às observações, assim, por essa transformação dos eixos, o método consegue retirar a correlação inicial das informações e manter apenas sua variância, isso nas  $n$ -dimensões para as quais se deseja resumir os dados (EFRON e TIBISHIRANI, cap. 5, 1993).

Segundo a descrição do programa:

[Principal component analysis (PCA) is a statistical technique for data reduction. The leading eigenvectors from the eigen decomposition of the correlation or covariance matrix of the variables describe a series of uncorrelated linear combinations of the variables that contain most of the variance. In addition to data reduction, the eigenvectors from a PCA are often inspected in order to learn more about the underlying structure of the data.]

O STATA fornece então os autovetores que são usados nos três componentes para simplificar cada uma das variáveis, que são depois somadas:

pca apvideo comp\_imp multimid, means

```
Principal components/correlation          Number of obs   =    2458
                                           Number of comp. =     3
                                           Trace           =     3
Rotation: (unrotated = principal)        Rho             =    1.0000
```

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	1.86748	1.1689	0.6225	0.6225
Comp2	.698578	.264635	0.2329	0.8554
Comp3	.433943	.	0.1446	1.0000

Principal components (eigenvectors)

Variable	Comp1	Comp2	Comp3	Unexplained
apvideo	0.6049	-0.3962	0.6907	0
comp_imp	0.6130	-0.3220	-0.7215	0
multimid	0.5083	0.8598	0.0481	0

Summary statistics of the variables

Variable	Mean	Std. Dev.	Min	Max
apvideo	4.976404	3.121253	0	41
comp_imp	5.614321	7.856942	0	71
multimid	1.353946	.9284686	0	5

```
gen pcavideo1 = 0.6049* apvideo
gen pcavideo2 = -0.3962* apvideo
gen pcavideo3 = 0.6907* apvideo

gen pcav = pcavideo1+ pcavideo2+ pcavideo3

gen pcacomp1 = 0.6130* comp_imp
gen pcacomp2 = -0.3220* comp_imp
gen pcacomp3 = -0.7215* comp_imp

gen pcac = pcacomp1+ pcacomp2+ pcacomp3

gen pcamult1 = 0.5083* multimid
gen pcamult2 = 0.8598* multimid
gen pcamult3 = 0.0481* multimid

gen pcam = pcamult1+ pcamult2+ pcamult3
```

Daí somamos os PCA's obtidos:

```
gen vcm2 = pcav+ pcac+ pcam
```

E podemos pedir para que tire valores negativos, somar16 no caso da 4ª série:

```
gen vcm3 = vcm2+16
```

Para a 8ª série o processo se repete com os autovalores obtidos para essa série sendo que, para evitar os negativos, somamos 16.5, e na 3ª série, com a mesma finalidade, somamos 5.

## **Anexo D**

No capítulo 3 vimos que os índices de eficiência obtidos pela DEA se dividem quanto à escala de rendimentos. Os dois principais são o de rendimentos constantes de escala, DEA-C, e o de rendimentos variáveis, DEA-V, que capta qualquer tipo de rendimento, inclusive os constantes. Observamos que o DEA-V é mais geral do que o DEA-C, já que todas as observações decretadas eficientes no DEA-C também o são para o DEA-V.

O motivo para se trabalhar com os dois índices reside no fato de que, quando os rendimentos são constantes, o cálculo e a convergência do DEA-C são mais rápidos, sendo portanto uma estimação mais eficiente do que o DEA-V. No entanto, quando os rendimentos são, de fato, variáveis, a obtenção dos índices de eficiência por DEA-C será viesada. Como ressaltam Simar e Wilson (2002, p. 116):

[Some authors have a priori imposed the rather restrictive assumption of constant returns to scale while using DEA methods; this may seriously distort measures of efficiency if the true technology displays non-constant returns of scale.]

É necessário, então, obter uma forma de verificação dos rendimentos de escala da fronteira. Uma maneira simples de se medir, é verificar a relação entre os índices DEA-C/DEA-V. Como o DEA-V é sempre maior ou igual ao índice DEA-C, um indicador constituído dessa forma possui limite superior igual a 1.

A idéia básica por trás da relação DEA-C/DEA-V é captar a distância entre os índices. Quando o DEA-V não se afasta muito do DEA-C, a densidade desse estimador estará concentrada próxima do limite superior 1, mostrando que a discrepância entre os índices é meramente estatística, não podendo ser, portanto, rejeitada a possibilidade de rendimentos decrescentes de escala. Para o caso contrário, a relação se mostrará afastada do limite superior, indicando um distanciamento do DEA-V, nessa situação, se torna forte a indicação para

rejeitarmos os rendimentos constantes de escala e adotarmos a suposição de rendimentos variáveis.

Para saber quando poderemos afirmar qual é o rendimento por trás da fronteira de eficiência é preciso testar se o processo gerador dos dados,  $P$ , indica realmente rendimentos constantes ou variáveis. Para isso, podemos construir o teste de hipóteses. A hipótese nula ( $H_0$ ) é de rendimentos constantes de escala (CRS), a hipótese alternativa ( $H_1$ ) é para o caso de rendimentos variáveis.

**$H_0$ :  $P$  é CRS.**

**$H_1$ :  $P$  é VRS.**

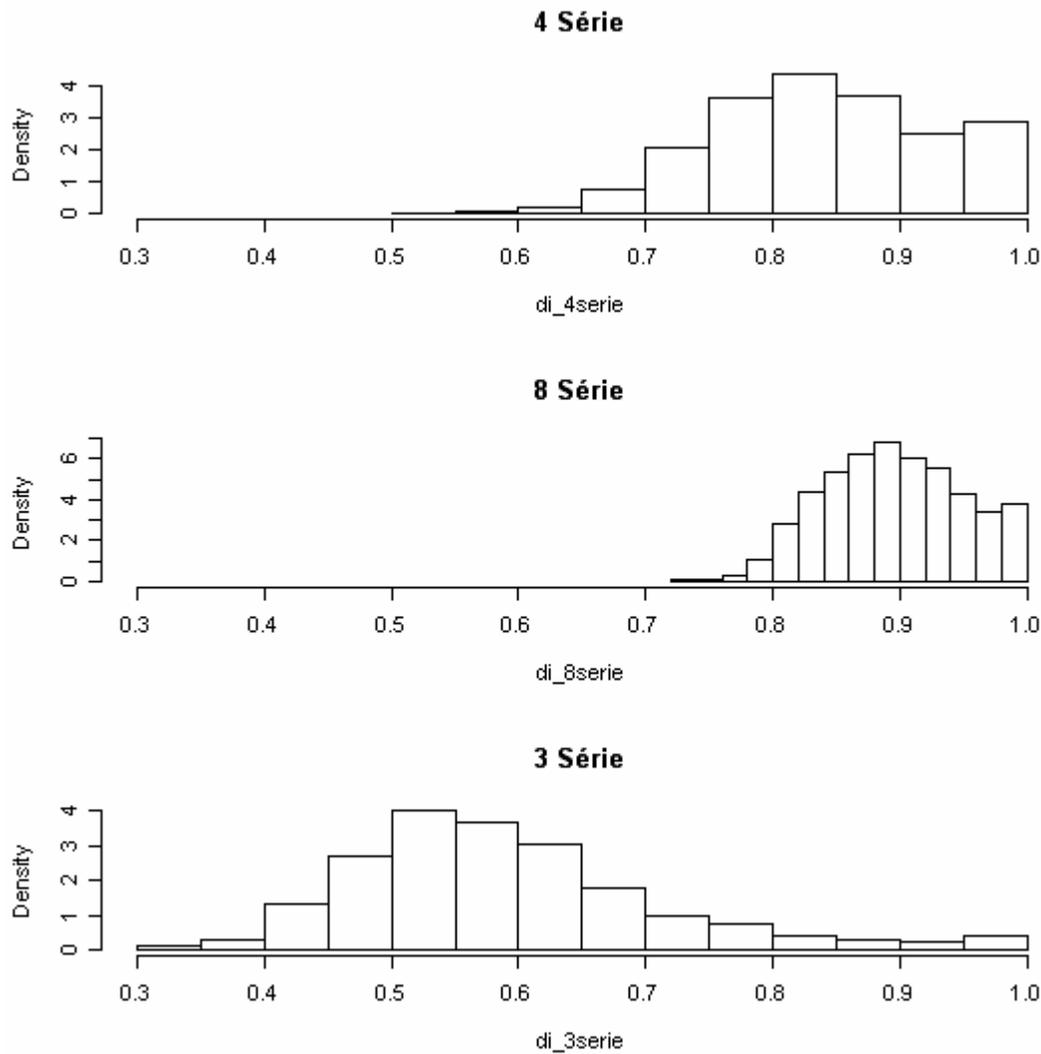
Uma das primeiras estatísticas que podemos construir para o teste de hipóteses acima é baseada na relação entre DEA-C e DEA-V, utilizando o princípio do distanciamento entre os dois índices. Os exemplos de testes que utilizam essa mesma idéia estão presentes em Byrnes et al (1986), Ferrier (1994), assim como também é revisitada em Simar e Wilson (2002, p. 119):

$$\hat{d}_i = \hat{\delta}_{i,DEA-C} / \hat{\delta}_{i,DEA-V}$$

O índice é construído para cada uma das observações  $i = 1, 2, \dots, n$ . No entanto, para que a estatística acima sirva para um teste de hipóteses, é preciso estabelecer um critério de “corte”, um d-crítico ( $d_c$ ) para o qual  $H_0$  possa ser rejeitada quando  $d_i < d_c$ . Conhecendo-se a distribuição estatística de  $d_i$  é possível obter um critério de corte para a rejeição de  $H_0$ . Simar e Wilson ressaltam que se supomos que cada realização do teste  $d_i$  é independente, o número de caso onde se rejeita a hipótese nula será uma distribuição binomial.

Antes de determinarmos o valor da estatística de corte para o d-crítico,  $d_c$ , apresentamos os histogramas da estatística  $\hat{d}_i$

**Gráfico D.1. Histogramas para estatística  $d_i$ , todas as séries.**



No gráfico acima já podemos antecipar que a 3ª série do ensino médio possui chances bem maiores de rejeitar a hipótese nula de retornos constantes de escala, pois observamos que a concentração de  $d_i$  está bem mais afastada de 1 do que nas

outras séries, indicando que DEA-V está distante do DEA-C. Sendo uma binomial, a estatística de teste será dada pela fórmula:

$$I_{\alpha}(r, n-r) = \sum_{j=r}^n \binom{n}{j} \alpha^j (1-\alpha)^{n-j}$$

O  $I_{\alpha}$  mede a probabilidade de  $r$ , número de rejeições, ser menor ou igual à  $n$  quando  $H_0$  é verdadeira. O  $\alpha$  é a probabilidade de ‘sucesso’, repare que o sucesso está designado pelo número de rejeições  $r$ , o  $j$  é o índice do somatório, igual a  $r$ , mas seu valor muda de forma a somar todos os números desejados. Dessa forma, um dos critérios de corte fica estabelecido quando impomos  $I_{0.05}(r, n-r) < 0.05$ . Ou seja, o número de rejeições tem de ser menor do 5% da amostra.

Se estabelecermos  $\alpha = 5\%$ , para a 4ª série temos de ter menos de **123** escolas abaixo do valor crítico para não rejeitar  $H_0$ .

Para a 8ª série: **127**;

Para a 3ª série do ensino médio: **71**.

Há um teste mais rigoroso que consiste em descobrir o valor de  $\alpha$  para que apenas uma rejeição invalide a hipótese de rendimentos decrescentes, este teste é aplicado a cada observação individualmente:

$$\Pr(r \geq 1) = 1 - \Pr(r = 0) = 1 - \binom{n}{0} \alpha^0 (1 - \alpha_l)^n = (1 - \alpha_l)^n$$

Temos de descobrir  $\alpha_l$ . Estabelecendo  $\Pr(r \geq 1)$  como  $(1 - \alpha_g)$  temos:

$$\alpha_l = 1 - (1 - \alpha_g)^{1/n}$$

Em nossa análise, onde as amostras são grandes, valores de  $n$  elevados nos trazem um  $\alpha_l$  muito pequeno, com  $\alpha_g$  de 0.05 temos:  $2.08e^{-5}$  para 4ª série;  $2.02e^{-5}$  para a 8ª, e  $3.61e^{-5}$ , valores muito baixos onde não chegamos a rejeitar  $H_0$  para nenhum dos casos.

Dois outros testes, propostos por Simar e Wilson, foram construídos:

O da média das distâncias:

$$\hat{S}_{1n} = n^{-1} \sum_{i=1}^n \hat{\delta}_{n,DEA-C}(x_i, y_i) / \hat{\delta}_{n,DEA-V}(x_i, y_i)$$

E uma variação deste último, a relação das médias:

$$\hat{S}_{2n} = \sum_{i=1}^n \hat{\delta}_{n,DEA-C}(x_i, y_i) / \sum_{i=1}^n \hat{\delta}_{n,DEA-V}(x_i, y_i)$$

Nota-se que as últimas estatísticas são valores agregados de todas as observações, portanto, seus intervalos foram construídos por *bootstrap*. Para as demais estatísticas utilizamos a distribuição do índice clássico, ou do índice *bootstrap*, indicado por *boot*. Simar e Wilson argumentam que é preciso tomar cuidado com a suposição de independência para o índice clássico, mas a amostra utilizada nesse trabalho possui tamanho considerável, não se tratando de problema muito grave no caso da relação DEA-C/DEA-V.

Os resultados dos valores críticos para cada um dos testes estão expressos na tabela abaixo, quando o teste rejeita a hipótese nula, o valor crítico está assinalado em negrito:

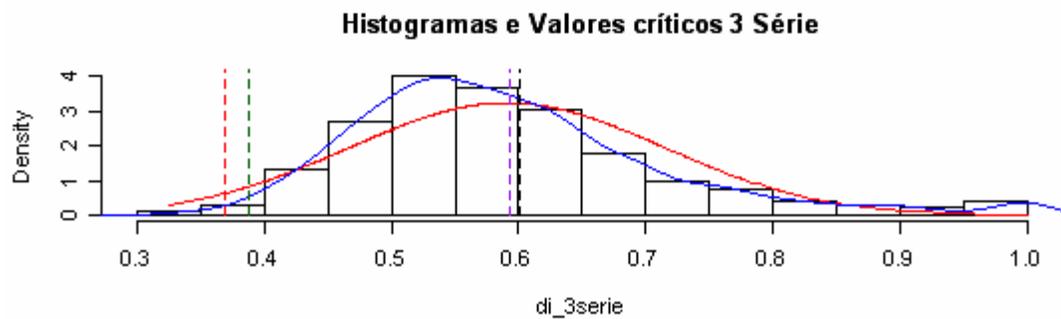
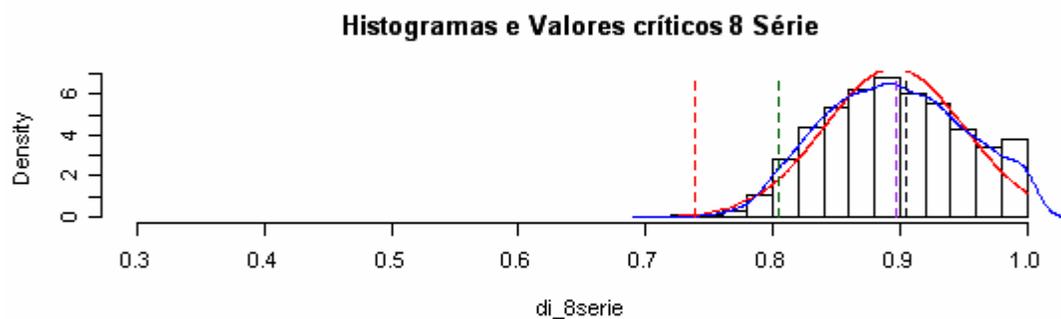
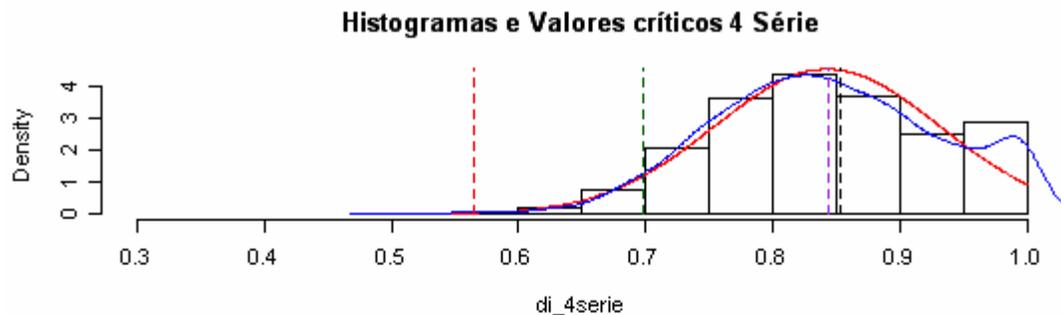
**Tabela D.1 Valores dos  $d_i$  críticos para os testes de escala.**

Teste	4ª Série			8ª Série			3ª Série		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
Binomial	0.566	0.542	0.523	0.749	0.735	0.724	0.392	0.359	0.332
Binomial Boot	0.605	0.565	0.534	0.769	0.739	0.729	<b>0.409</b>	<b>0.369</b>	<b>0.337</b>
Normal Binomial	0.595	0.537	0.428	0.595	0.537	0.428	<b>0.595</b>	<b>0.537</b>	<b>0.428</b>
Normal	0.729	0.697	0.637	<b>0.825</b>	0.805	0.767	0.433	0.388	0.304
Normal Boot	<b>0.725</b>	<b>0.680</b>	0.595	<b>0.826</b>	<b>0.799</b>	0.748	<b>0.431</b>	<b>0.381</b>	<b>0.287</b>
S1n	0.843	0.843	0.843	0.895	0.895	0.895	<b>0.591</b>	<b>0.591</b>	<b>0.591</b>
S2n	0.841	0.840	0.839	0.895	0.894	0.894	0.588	0.587	0.584

Os resultados acima reforçam a opção de usar retornos variáveis para a 3ª série. Abaixo apresentamos os resultados gráficos dos valores críticos a 5%, em vermelho está a estatística Binomial, de verde a o critério de corte a partir da Normal com média e desvio de acordo com a distribuição, a de roxo é a estatística  $S_{2n}$ , muito próxima da de preto que é a  $S_{1n}$ . A aplicação de  $S_{1n}$  e  $S_{2n}$ , precisa de um intervalo construído por *bootstrap* para saber se o valor da estatística está fora da distribuição. A regra aqui adotada foi um tanto mais simples: calculamos os indicadores de  $S_{1n}$  e  $S_{2n}$  aplicado aos índices  $\delta^*_{DEA-C}/\delta^*_{DEA-V}$ , depois verificamos se este resultado era menor do que a estatística comum, sem o *bootstrap*, daí sim, construímos os intervalos da estatística por meio de *bootstrap* simples.<sup>67</sup>

<sup>67</sup> Para pequenas amostras é preciso se aplicar o *bootstrap-corrigido*, no entanto a grande amostra trabalhada dá liberdade para construirmos um intervalo menos rigoroso.

**Gráfico D.2. Histogramas da distribuição dos índices de eficiência com os respectivos cortes.**



O índice para a detecção de *outliers* de Sampaio de Sousa e Stosic (2005) possui uma série de passos que visam abreviar e poupar o esforço computacional de retirar as observações uma a uma, os passos da construção do Sampaio-Stosic visam a aplicar um processo estocástico mais eficiente combinando a técnica de *bootstrap* com o *jackknife*, tais passos serão aqui repetidos conforme as páginas 163-164 no artigo de 2005 destes referidos autores:

Relembremos que a estatística *leverage* era dada pela equação (18):

$$l_j = \sqrt{\frac{\sum_{k=1; k \neq j}^k (\delta_{kj}^* - \delta_k)^2}{K-1}}$$

- [1] Selecionar um subconjunto aleatório de  $L$  observações dentro do espaço amostral que inclui todas as  $K$  DMU's, (os autores denominam essa sub-amostra de “bolha” onde o tamanho de  $L$  pode variar de 10-20% da amostra). A partir desse processo, obter os *leverages*,  $l_j$ , de cada uma das estatísticas escolhidas de dentro do conjunto  $\{1, \dots, K\}$ .
- [2] O passo acima é repetido  $B$  vezes, acumulando as informações de várias “bolhas” de *leverage*,  $\tilde{l}_{jb}$ . O  $B$  deve ser grande o suficiente para que cada observação tenha a chance de ser escolhida aproximadamente um número  $n_j \approx BL/K$  de vezes.
- [3] Calcula-se o *leverage* médio para cada DMU escolhida:

$$\tilde{l}_j = \frac{\sum_{b=1}^{n_j} \tilde{l}_{jb}}{n_j}$$

E a média global do *leverage*:

$$\tilde{l} = \frac{\sum_{j=1}^K \tilde{l}_j}{K}$$

- [4] O *bootstrap* desse processo de re-amostragem é realizado sem reposição o que, para fins de compor uma estatística de *leverage*, é mais útil, já que o propósito no trabalho de Sampaio de Sousa e Stosic, assim como o nosso, não é compensar um pequeno tamanho da base de dados, mas sim tornar factível o computo de planilhas de dados muito grandes.

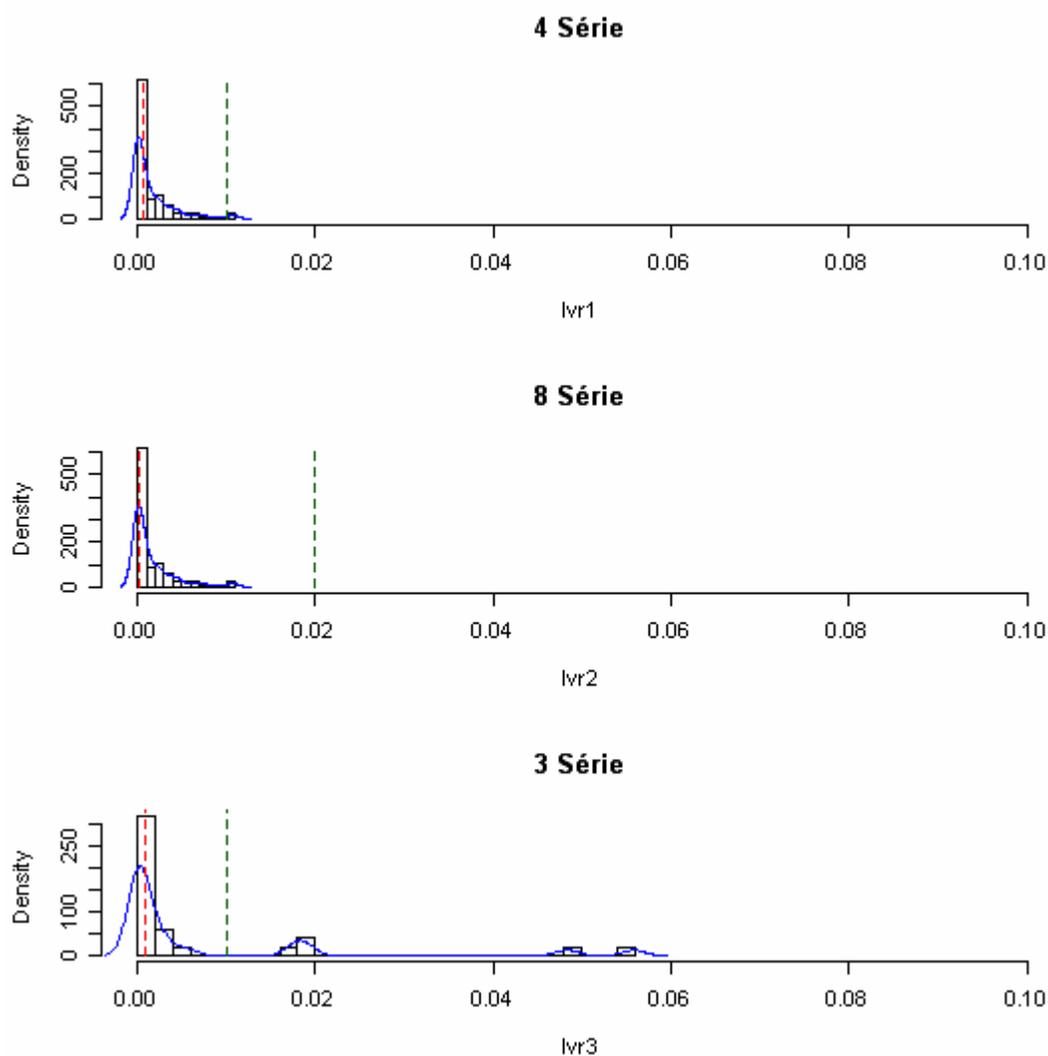
A base de dados dessa dissertação é um pouco menor do que a presente em Sampaio-Stosic, nossa maior série possui 2431 observações, enquanto que a base dos autores mencionados possuía 4796 municípios. A base um pouco menor e sub-dividida por série nos permitiu calcular o *leverage* de duas maneiras: uma de acordo com o processo acima mencionado e outra de maneira mais simples, porém mais intensiva em tempo computacional, que consiste aplicar o *leverage* direto na amostra, retirando as observações uma por vez. Esse último método foi o efetivamente empregado na dissertação, mas os resultados de ambos são semelhantes.

Para determinarmos se a observação é ou não influente, precisamos de um critério de corte. Sampaio de Sousa et al. sugerem como critério de corte algum múltiplo da média global do *leverage*,  $\tilde{l}_0 = c\tilde{l}$ , onde  $c$  é, via de regra, 2 ou 3. Para os dados de escolas estaduais de Minas Gerais, a média global do *leverage* possui um nível baixo, e todas as informações “dentro” da fronteira possuem em  $l_0 = 0$ , ou seja, muitas das informações não exercem nenhuma influência na fronteira. Para cada uma das séries foram utilizados dois valores de corte para o *leverage*. O primeiro, mais rigoroso capta apenas a ponta final da distribuição de *leverages*, o segundo varia para captar cerca de 2,5% da cauda, estes valores de corte estão especificados na tabela D.2 abaixo:

**Tabela D.2 Valores dos  $l_c$  críticos para teste de detecção de outliers.**

Série	Critério 1	Outliers critério 1	Critério 2	Outliers critério 2
4 <sup>a</sup> série	0.01	2	6.66e-4	35
8 <sup>a</sup> série	0.01	1	1.00e-5	51
3 <sup>a</sup> série	0.02	2	9.40e-5	17

**Gráfico D.3. Histogramas da distribuição dos leverages com respectivos cortes.**



O índice Sampaio-Stosic possui a desvantagem de perder informação a partir do descarte de observações, no entanto, isso não se constitui problema muito grave, pois, geralmente, são poucos as informações descartadas, além disso, o pesquisador pode identificá-las individualmente e considerar se aquela informação é *outlier* por discrepância estatística, ou se é realmente algum nível de excelência no padrão de uso de recursos, insumos e produtos.

Como forma de identificar a eficácia do índice S-S e dos demais *rankings* de escala gerados nesse estudo, testamos a aderência entre o ranking das escolas presentes em nossa base com uma lista de 220 escolas consideradas de qualidade pela Secretaria de Educação do Estado.<sup>68</sup>

A aderência foi testada através da correlação entre os *rankings*, se estava presente tanto no *ranking* da Secretaria quanto no *ranking* das 250 melhores no índice de eficiência, ganhava escore 1, caso contrário 0. Foi testado também o *ranking* entre as escolas de qualidade e o *ranking* apenas com a eficiente de fronteira,  $\delta_i = 1.000$  e o *ranking* apenas se pertence à lista da Secretaria com relação direta ao índice de eficiência dos 250 primeiros. Os resultados estão representados na tabela abaixo.

**Tabela D.3 Teste de Aderência entre escolas de Qualidade e a eficiência DEA.\***

	$\delta_v = 1.000$	$\delta_{v250}$	$\delta_{02}$	$\delta_{02-250}$	$\delta_v^*$
<b>Correlação entre 0 e 1</b>	0.089	0.121	0.112	0.170	0.110
<b>Correlação direta com o índice</b>	0.161		0.191		0.168

\* Lembrar que para a 3ª série todas as eficiências são referente à DEA-V, inclusive  $\delta_{02}$ .

A correlação de 0.17 para o  $\delta_{02-250}$  indica que das 220 escolas de eficiência da Secretaria, 68 eram as mesmas escolas no ranking S-S, ou seja, a eficiência construída sem os *outliers* identificados por S-S, reconhece como eficiente 30% das

<sup>68</sup> No critério da secretaria a escolha parece se basear muito nas escolas que ministram a 3ª série do ensino médio, não é exclusivo, há algumas escolas de 4ª série e outras de 8ª, mas o ensino médio comparece em maior peso. A aderência foi verificada para o caso da 3ª série portanto.

escolas da Secretaria.<sup>69</sup> O índice Sampaio-Stosic é o melhor para captar não só *outliers* como também é o mais preciso na identificação de eficiência. A eficiência *bootstrap* também corrige os *outliers*, no entanto, como o propósito deste último índice é verificar o intervalo de probabilidade da eficiência e não a identificação da observação em si, seu desempenho na correlação é o mais fraco.<sup>70</sup>

O teste CFS se aplica tanto aos métodos não-paramétricos quanto aos paramétricos de fronteiras estocásticas. Forneceremos aqui sua intuição inicial aplicada ao caso um-*input* e um-*output*, no entanto, o teste se aplica também aos casos de múltiplos *inputs* e *outputs*, como explica Simar (2003, pp. 399).

Imaginemos uma distribuição conjunta de *inputs* e *outputs*. Para a análise de eficiência, podemos nos perguntar sobre qual a probabilidade de uma observação utilizar um nível de insumos (**X**) maior do que  $x$ , dado que impusemos a ela uma produção **Y** maior do que  $y$ :

$$S_c(x|y) = \text{Prob}(X \geq x | Y \geq y) = \frac{S(x, y)}{S_y(y)} \quad (21)$$

Onde  $S(x, y)$  é uma distribuição de  $\text{Prob}(X \geq x | Y \geq y)$  e  $S_y(y) = \text{Prob}(Y \geq y)$  que é a probabilidade marginal quando  $x = 0$  em  $S(x, y)$ . A partir de  $S_c(x|y)$  definimos um limite inferior para esta função que lembra as funções de sobrevivência:

$$\phi(y) = \min \{ x | S_c(x|y) < 1 \} \quad (22)$$

---

<sup>69</sup> A correlação é um pouco menor do que esses 30% porque há a contagem da relação no sentido contrário, as que foram declaradas eficientes no  $\delta_{o2}$  mas não na lista, ou vice-versa.

<sup>70</sup> Por ser mais suave, a *DEA-bootstrap* é o melhor critério para ser usada no 2º estágio, no entanto, como se evidencia nos modelos das regressões tobit (6), onde é empregado a DEA-clássica, o sentido das regressões não destoa muito.

Como não temos as funções reais, trabalhamos com as observadas. Lançando mão do princípio *plug-in* (Simar 2003, pp 396):

$$\hat{S}_{c,n}(x|y) = \frac{\hat{S}_n(x, y)}{S_{yn}(y)} \quad (23)$$

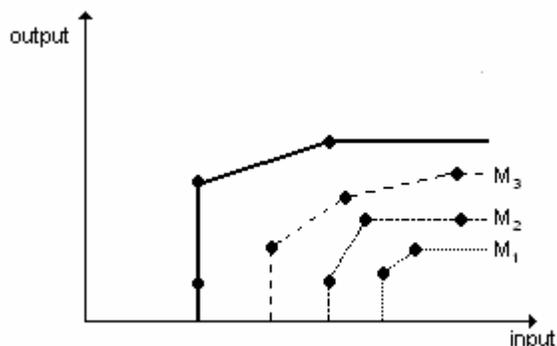
$$\hat{\phi}_n(y) = \min\{x | \hat{S}_{c,n}(x|y) < 1\} \quad (24)$$

A partir de então a idéia do teste CFS é obter vários limites inferiores a partir de ‘m’ amostras da planilha inicial que contenham X maior ou igual do que o ponto observado x ( $X \geq x$ ) e Y maiores do que o observado y ( $Y \geq y$ ). Temos:

$$\phi_m = E[\min(X^1, \dots, X^m)] = \int_0^\infty [S(x)]^m dx \quad (25)$$

Em uma representação gráfica teríamos:

**Gráfico D.4. Exemplos de Fronteiras Interniores - modelo CFS de exclusão de *outliers*.**



É possível provar que para  $\lim_{m \rightarrow \infty}$  a fronteira de limite mínimo tenderá à fronteira verdadeira  $\phi$ . Os  $m$ 's são re-amostragens aleatórias, portanto, os  $M_1$ ,  $M_2$  e  $M_3$  do

gráfico não precisam estar ordenados necessariamente daquela forma, estão a título de exemplo. A intenção é mostrar que quanto maior a amostra, maior a probabilidade de convergir para a verdadeira fronteira que, a princípio, não é possível saber onde se situa. No entanto, o que o teste de detecção de *outliers* CFS acusa é que caso 'm' se expanda e os pontos 1, 2 e 3 continuem sempre (ou a maioria das vezes) fora das novas fronteiras obtidas, é forte o indício de serem outliers.

Em termos gerais este é o princípio do teste CFS. Cazals et al. (2002) mostram que essa forma de medição converge para o *Free Disposal Hull* – FDH, que é o método DEA livre da convexidade, ou seja, é um teste apropriado para o caso não paramétrico e, além disso, nada o impede de ser um teste de detecção de *outliers* também no caso paramétrico.

## Anexo E

As regressões (1) e (2) são mais criteriosas e servem de guia para as demais por envolverem o seguinte algoritmo proposto por Simar e Wilson (2003, p.13):

### # Algoritmo Simar e Wilson:

- [1] A partir dos dados originais  $L_N = \{(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i, \mathbf{d}_i)\}$ , calcular  $\hat{\delta}_i = \hat{\delta}(x_i, y_i | \hat{P})$ ,  $\forall i=1, \dots, n$  usando as equações (5)-(6) e a equação (8) para o caso da 3ª série.
- [2] O método de MV é usado para obter estimações de  $\hat{\beta}$  do  $\beta$ -verdadeiro, assim como a obtenção de  $\hat{\sigma}_\varepsilon$  do  $\sigma_\varepsilon$ -verdadeiro na regressão truncada de  $\delta_i$  em  $\mathbf{z}_i$  presente em (16).
- [3] Repetir os passos ([3.1]-[3.3])  $B$  vezes para obter um conjunto de estimadores *bootstrap*:  $A = \{(\hat{\beta}^*, \hat{\sigma}_\varepsilon^*)\}_{b=1}^B$ .
  - [3.1] Para cada  $i = 1, \dots, n$  obter os erros  $\varepsilon_i$  a partir de  $N(0, \sigma_\varepsilon)$  truncada à esquerda em  $(1 - z_i \hat{\beta})$ .
  - [3.2] Novamente para cada  $i = 1, \dots, n$  calcular  $\hat{\delta}_i^* = z_i \hat{\beta} + \varepsilon_i$
  - [3.3] Usar MV para estimar a regressão truncada de  $\delta_i^*$  em  $\mathbf{z}_i$  obtendo as estimativas  $(\hat{\beta}, \hat{\sigma}_\varepsilon)$ .
- [4] Usar os valores *bootstrap* de  $A$  e as estimativas originais de  $\hat{\beta}, \hat{\sigma}_\varepsilon$  para construir os intervalos de confiança de  $\beta$  e  $\sigma_\varepsilon$ .

Inserimos também neste Anexo as Tabelas do Capítulo 5 que foram omitidas com objetivo de economizar espaço:

**Tabela E.1. Resumo para Variáveis de Atendimento e Infra-estrutura Eficientes 4ª série:**

Eficientes		Não Eficientes	
Variável		Variável	
urbana	0.77	urbana	0.80
fnde	0.77	fnde	0.82
liv4serie	0.23	liv4serie	0.26
liv8serie	0.34	liv8serie	0.34
merenda	0.95	merenda	0.96
TemTV	0.89	TemTV	0.98
TemPC	0.63	TemPC	0.50
TemMult	0.74	TemMult	0.91
sani_den	0.94	sani_den	0.94
ener_pub	0.97	ener_pub	0.98
agua_pub	0.79	agua_pub	0.84
esg_pub	0.67	esg_pub	0.69
lixo_col	0.73	lixo_col	0.78

Fonte: Censo Escolar MEC/IBGE 2003.

**Tabela E.2. Resumo para Variáveis de Dotação 4 Série:**

Eficientes		Não Eficientes	
Variável		Variável	
cor	0.43	cor	0.41
esc_resp_fem	0.35	esc_resp_fem	0.30
livros	0.24	livros	0.21
computador	0.15	computador	0.12
ler_livro_inf	0.80	ler_livro_inf	0.77
ler_jornais	0.47	ler_jornais	0.43
trab_fora	0.85	trab_fora	0.84
trab_domestico	0.29	trab_domestico	0.28
devermat	0.72	devermat	0.71
nreprovado	0.75	nreprovado	0.73

Fonte: Questionário SIMAVE 2003.

**Tabela E.3. Média dos Produtos e Insumos por Mesoregião 8ª série.**

Código	Mesoregião	#	Produtos				Insumos			
			<i>Pmate</i>	<i>Pport</i>	<i>Alunos</i>	<i>Custo</i>	<i>Qdoc</i>	<i>Salas</i>	<i>Infra</i>	<i>Vcm</i>
310001	Noroeste de Minas	60	235.32	236.65	368.03	53.03	12.25	21.28	4.63	16.00
310002	Norte de Minas	347	218.58	230.01	381.61	37.41	10.21	17.56	3.52	16.60
310003	Jequitinhonha	163	229.74	232.68	401.52	45.60	8.57	17.45	4.01	16.17
310004	Vale do Mucuri	71	220.99	229.68	411.73	50.50	10.46	19.27	4.21	15.93
310005	Triângulo Mineiro	226	245.42	244.89	441.34	60.06	16.30	23.82	4.82	15.88
310006	Central Mineira	62	242.81	243.70	408.37	57.35	13.90	20.32	4.53	16.05
310007	Metropolitana	557	239.52	241.95	478.98	48.07	15.64	24.26	4.77	15.75
310008	Vale do Rio Doce	307	235.77	237.18	332.90	56.95	11.05	17.35	4.43	16.42
310009	Oeste de Minas	117	248.02	246.80	442.99	62.03	15.94	22.05	4.79	15.92
310010	Sul/Sudoeste de Minas	259	250.10	245.70	483.98	58.48	16.85	21.02	4.90	15.82
310011	Campo das Vertentes	70	244.90	242.90	402.74	61.70	16.26	19.79	4.83	15.54
310012	Zona da Mata	292	242.22	244.10	391.70	58.73	14.62	20.07	4.83	15.89
<b>310013</b>	<b>Minas Gerais</b>	<b>2531</b>	<b>237.49</b>	<b>239.85</b>	<b>419.99</b>	<b>55.82</b>	<b>13.71</b>	<b>20.68</b>	<b>4.51</b>	<b>16.03</b>

**Tabela E.4. Distribuição das escalas de rendimento por Mesoregião 8ª série:**

Código	Mesoregião	#	<i>Constante</i>	<i>Decrecente</i>	<i>Crescente</i>
310001	Noroeste de Minas	60	0.00%	100.0%	0.00%
310002	Norte de Minas	347	4.61%	93.08%	2.31%
310003	Jequitinhonha	163	3.07%	95.09%	1.84%
310004	Vale do Mucuri	71	1.41%	98.59%	0.00%
310005	Triângulo Mineiro	226	1.77%	97.79%	0.44%
310006	Central Mineira	62	3.23%	96.77%	0.00%
310007	Metropolitana	557	2.69%	96.59%	0.72%
310008	Vale do Rio Doce	307	2.61%	97.07%	0.33%
310009	Oeste de Minas	117	1.71%	98.29%	0.00%
310010	Sul/Sudoeste de Minas	259	0.77%	99.23%	0.00%
310011	Campo das Vertentes	70	2.86%	97.14%	0.00%
310012	Zona da Mata	292	1.71%	97.95%	0.34%
<b>310013</b>	<b>Minas Gerais</b>	<b>2531</b>	<b>2.45%</b>	<b>96.84%</b>	<b>0.71%</b>

**Tabela E.5. Resumo Variáveis da Função para Escolas Eficientes 8 série:**

	<b>Média</b>	<b>Desvio</b>	<b>Min</b>	<b>Max</b>
<b>Matemática</b>	247.03	21.30	184.56	315.56
<b>Português</b>	246.01	13.12	165.30	281.46
<b>Matrículas 4 série</b>	461.40	300.63	8.00	1396.00
<b>Matrículas Escola</b>	1458.60	919.10	108.00	6126.00
<b>Custo Dentro</b>	55.89	19.89	9.20	209.12
<b>Docentes com superior</b>	12.06	10.95	0.00	52.00
<b>Salas</b>	17.06	8.98	2.00	66.00
<b>Infraestrutura</b>	4.21	1.15	1.00	5.00
<b>VCM</b>	15.32	2.04	7.03	18.09

**Tabela E.6. Resumo Variáveis da Função para Escolas NÃO Eficientes 8 série:**

	<b>Média</b>	<b>Desvio</b>	<b>Min</b>	<b>Max</b>
<b>Matemática</b>	233.38	18.33	164.00	306.79
<b>Português</b>	237.21	13.38	164.06	296.77
<b>Matrículas 4 série</b>	402.07	226.31	32.00	1406.00
<b>Matrículas Escola</b>	1355.99	726.91	114.00	4048.00
<b>Custo Dentro</b>	55.79	18.01	6.27	203.49
<b>Docentes com superior</b>	14.41	11.30	0.00	108.00
<b>Salas</b>	22.22	10.46	2.00	136.00
<b>Infraestrutura</b>	4.64	0.78	0.00	5.00
<b>VCM</b>	16.34	1.63	0.13	19.44

**Tabela E.7. Resumo para Variáveis de Atendimento e Infraestrutura Eficientes 8ª :**

Eficientes		Não Eficientes	
Variável		Variável	
urbana	0.82	urbana	0.92
fnde	0.76	fnde	0.80
liv4serie	0.18	liv4serie	0.21
liv8serie	0.41	liv8serie	0.42
merenda	0.93	merenda	0.96
TemTV	0.98	TemTV	0.99
TemPC	0.71	TemPC	0.66
TemMult	0.83	TemMult	0.92
sani_den	0.94	sani_den	0.97
ener_pub	1.00	ener_pub	1.00
agua_pub	0.83	agua_pub	0.94
esg_pub	0.69	esg_pub	0.82
lixo_col	0.76	lixo_col	0.90

**Tabela E.8. Resumo Variáveis Dotação 8 Série:**

Eficientes		Não Eficientes	
Variável		Variável	
cor	0.49	cor	0.44
esc_resp_fem	0.21	esc_resp_fem	0.18
livros	0.26	livros	0.24
computador	0.12	computador	0.10
ler_livro_inf	0.73	ler_livro_inf	0.69
ler_jornais	0.45	ler_jornais	0.45
trab_fora	0.65	trab_fora	0.63
trab_domestico	0.26	trab_domestico	0.26
devermat	0.57	devermat	0.54
nreprovado	0.69	nreprovado	0.66

**Tabela E.9. Média dos Produtos e Insumos por Mesoregião 3ª série.**

Código	Mesoregião	#	Produtos				Insumos			
			<i>Pmate</i>	<i>Pport</i>	<i>Alunos</i>	<i>Custo</i>	<i>Qdoc</i>	<i>Salas</i>	<i>Infra</i>	<i>Vcm</i>
310001	Noroeste de Minas	34	260.79	264.41	488.21	41.53	9.18	20.82	4.76	13.79
310002	Norte de Minas	138	250.05	260.10	535.40	25.11	7.57	19.78	3.89	13.27
310003	Jequitinhonha	62	264.93	267.13	446.90	33.17	6.16	19.89	4.47	12.06
310004	Vale do Mucuri	32	257.86	263.47	441.94	36.44	5.41	21.00	4.38	12.51
310005	Triângulo Mineiro	136	278.49	272.58	591.60	50.15	12.20	23.99	4.91	14.94
310006	Central Mineira	39	276.81	275.06	420.03	48.02	8.28	18.64	4.36	13.23
310007	Metropolitana	352	266.62	268.17	675.93	33.92	15.07	24.27	4.79	13.20
310008	Vale do Rio Doce	146	268.66	267.85	442.45	39.94	8.38	18.92	4.80	13.70
310009	Oeste de Minas	74	276.24	275.90	496.93	50.38	15.07	22.85	4.85	13.80
310010	Sul/Sudoeste de Minas	169	281.28	276.93	488.74	52.14	10.98	20.72	4.95	14.83
310011	Campo das Vertentes	53	277.04	275.67	416.00	50.68	10.77	19.36	4.91	14.50
310012	Zona da Mata	184	275.95	271.77	385.81	45.74	10.84	19.64	4.91	13.43
<b>310013</b>	<b>Minas Gerais</b>	<b>1419</b>	<b>270.07</b>	<b>269.92</b>	<b>524.39</b>	<b>41.02</b>	<b>11.25</b>	<b>21.47</b>	<b>4.72</b>	<b>13.68</b>

**Tabela E.10. Distribuição das escalas de rendimento por Mesoregião 3ª série:**

Código	Mesoregião	#	Constante	Decrecente	Crescente
310001	Noroeste de Minas	34	0.00%	100.00%	0.00%
310002	Norte de Minas	138	7.25%	91.30%	1.45%
310003	Jequitinhonha	62	1.61%	98.39%	0.00%
310004	Vale do Mucuri	32	0.00%	100.00%	0.00%
310005	Triângulo Mineiro	136	0.74%	99.26%	0.00%
310006	Central Mineira	39	2.56%	97.44%	0.00%
310007	Metropolitana	352	1.70%	98.01%	0.28%
310008	Vale do Rio Doce	146	0.68%	99.32%	0.00%
310009	Oeste de Minas	74	0.00%	100.00%	0.00%
310010	Sul/Sudoeste de Minas	169	0.00%	100.00%	0.00%
310011	Campo das Vertentes	53	0.00%	100.00%	0.00%
310012	Zona da Mata	184	0.54%	99.46%	0.00%
<b>310013</b>	<b>Minas Gerais</b>	<b>1419</b>	<b>1.48%</b>	<b>98.31%</b>	<b>0.21%</b>

**Tabela E.11. Resumo Variáveis da Função para Escolas Eficientes 3 série:**

	<b>Média</b>	<b>Desvio</b>	<b>Min</b>	<b>Max</b>
<b>Matemática</b>	247.03	21.30	184.56	315.56
<b>Português</b>	246.01	13.12	165.30	281.46
<b>Matrículas 4 série</b>	461.40	300.63	8.00	1396.00
<b>Matrículas Escola</b>	1458.60	919.10	108.00	6126.00
<b>Custo Dentro</b>	55.89	19.89	9.20	209.12
<b>Docentes com superior</b>	12.06	10.95	0.00	52.00
<b>Salas</b>	17.06	8.98	2.00	66.00
<b>Infraestrutura</b>	4.21	1.15	1.00	5.00
<b>VCM</b>	15.32	2.04	7.03	18.09

**Tabela E.12. Resumo Variáveis da Função para Escolas NÃO Eficientes 3 série:**

	<b>Média</b>	<b>Desvio</b>	<b>Min</b>	<b>Max</b>
<b>Matemática</b>	233.38	18.33	164.00	306.79
<b>Português</b>	237.21	13.38	164.06	296.77
<b>Matrículas 4 série</b>	402.07	226.31	32.00	1406.00
<b>Matrículas Escola</b>	1355.99	726.91	114.00	4048.00
<b>Custo Dentro</b>	55.79	18.01	6.27	203.49
<b>Docentes com superior</b>	14.41	11.30	0.00	108.00
<b>Salas</b>	22.22	10.46	2.00	136.00
<b>Infraestrutura</b>	4.64	0.78	0.00	5.00
<b>VCM</b>	16.34	1.63	0.13	19.44

**Tabela E.13. Resumo para Variáveis de Atendimento e Infraestrutura Eficientes 3ª série.**

Eficientes		Não Eficientes	
Variável		Variável	
urbana	0.99	urbana	0.99
fnde	0.83	fnde	0.73
liv4serie	0.21	liv4serie	0.19
liv8serie	0.40	liv8serie	0.43
merenda	0.93	merenda	0.94
TemTV	1.00	TemTV	0.99
TemPC	0.71	TemPC	0.77
TemMult	0.92	TemMult	0.95
sani_den	0.98	sani_den	0.98
ener_pub	1.00	ener_pub	1.00
agua_pub	0.95	agua_pub	0.97
esg_pub	0.80	esg_pub	0.92
lixo_col	0.91	lixo_col	0.96

**Tabela E.14. Resumo Variáveis Dotação 3ª série.**

Eficientes		Não Eficientes	
Variável		Variável	
cor	0.53	cor	0.52
esc_resp_fem	0.23	esc_resp_fem	0.20
livros	0.25	livros	0.28
computador	0.10	computador	0.12
ler_livro_inf	0.75	ler_livro_inf	0.70
ler_jornais	0.61	ler_jornais	0.64
trab_fora	0.48	trab_fora	0.47
trab_domestico	0.26	trab_domestico	0.26
devermat	0.46	devermat	0.40
nreprovado	0.65	nreprovado	0.62
noite	0.63	noite	0.70

Regressões com o indicador de presença ou ausência do quesito de infra-estrutura. Verificamos apenas a existência ou ausência de energia elétrica, água pública, esgotamento e coleta de lixo (no capítulo 5 a diferenciação era entre pública ou alguma maneira alternativa ou inexistente). O regressando é o inverso do índice DEA-C\* para 4ª e 8ª séries e o inverso da DEA-V\* para 3ª série do ensino médio.

**Tabela E.15. Regressões truncadas para novos indicadores de infraestrutura 4ª Serie.**

```
Truncated regression
Limit:  lower =      1
        upper =    +inf
Log likelihood = 620.18291
Number of obs = 2460
Wald chi2(9) = 1121.40
Prob > chi2 = 0.0000
```

dea_s	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----						
eql						
urbana	-.0702375	.011407	-6.16	0.000	-.0925947	-.0478802
apvideo	.0447703	.0015948	28.07	0.000	.0416446	.0478959
comp_imp	-.0162156	.0006481	-25.02	0.000	-.0174859	-.0149453
multimid	.0643409	.004872	13.21	0.000	.054792	.0738899
sani_den	-.0276069	.0175309	-1.57	0.115	-.0619669	.0067531
eletric	-.0927668	.030752	-3.02	0.003	-.1530396	-.032494
agua	.016145	.0225852	0.71	0.475	-.0281212	.0604112
esgoto	.0145487	.0282346	0.52	0.606	-.0407901	.0698875
lixo	.0022045	.0129821	0.17	0.865	-.0232399	.027649
_cons	1.524589	.0453126	33.65	0.000	1.435778	1.6134
-----						
sigma						
_cons	.1910682	.0028425	67.22	0.000	.185497	.1966394
-----						

**Tabela E.16. Regressões truncadas para novos indicadores de infraestrutura 8ª Serie.**

```
Truncated regression
Limit:  lower =      1
        upper =    +inf
Log likelihood = 2137.7089
Number of obs = 2533
Wald chi2(9) = 672.00
Prob > chi2 = 0.0000
```

dea_s	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----						
eql						
urbana	.0554114	.0073672	7.52	0.000	.040972	.0698508
apvideo	.0117089	.0008346	14.03	0.000	.0100732	.0133447
comp_imp	-.0080279	.0003629	-22.12	0.000	-.0087393	-.0073166
multimid	.0299516	.0028136	10.65	0.000	.024437	.0354662
sani_den	.0344448	.0120714	2.85	0.004	.0107853	.0581042
eletric	-.0403367	.0348742	-1.16	0.247	-.1086888	.0280154
agua	.017638	.0123702	1.43	0.154	-.0066071	.0418831
esgoto	.0778684	.0191579	4.06	0.000	.0403197	.1154171
lixo	.0102052	.0069038	1.48	0.139	-.003326	.0237365
_cons	1.126942	.0427827	26.34	0.000	1.043089	1.210794
-----						
sigma						
_cons	.1053211	.0015351	68.61	0.000	.1023125	.1083298
-----						

**Tabela E.17. Regressões truncadas para novos indicadores de infraestrutura 3ª Serie.**

Truncated regression  
 Limit: lower = 1 Number of obs = 1420  
 upper = +inf Wald chi2(9) = 117.08  
 Log likelihood = 2062.5142 Prob > chi2 = 0.0000

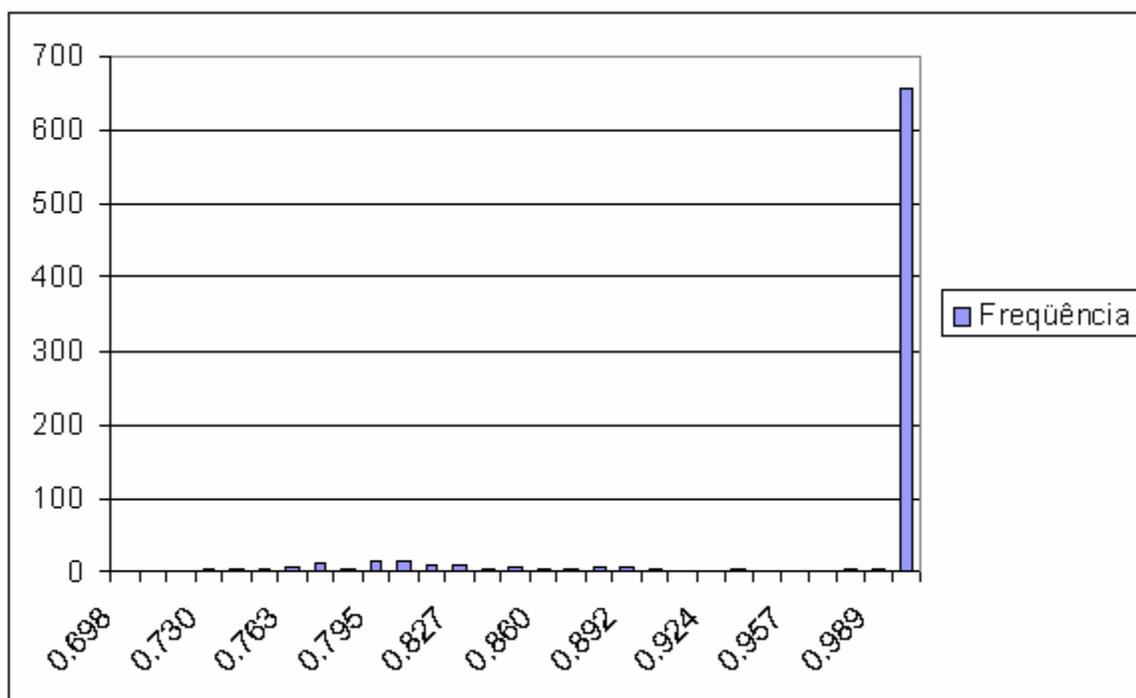
dea_s	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----						
eq1						
urbana	.0040044	.0161284	0.25	0.804	-.0276068	.0356156
apvideo	.0043337	.000573	7.56	0.000	.0032107	.0054568
comp_imp	-.0000486	.0002573	-0.19	0.850	-.000553	.0004558
multimid	.008382	.0021484	3.90	0.000	.0041711	.0125928
sani_den	.0019364	.0117311	0.17	0.869	-.0210561	.024929
eletric	-.0313219	.0368851	-0.85	0.396	-.1036154	.0409717
agua	-.0048328	.0096576	-0.50	0.617	-.0237614	.0140958
esgoto	.0026712	.0168454	0.16	0.874	-.0303451	.0356876
lixo	.0050018	.0053842	0.93	0.353	-.0055511	.0155547
_cons	1.118966	.0460773	24.28	0.000	1.028656	1.209276
-----						
sigma						
_cons	.059578	.0012828	46.44	0.000	.0570637	.0620923
-----						

## Anexo F

O objetivo deste anexo é demonstrar como foi observada a existência de dois grupos distintos entre os municípios da DEA-eficiência. Em um primeiro momento, realizando os testes para verificar a escala similares aos presentes no Anexo D, verificamos que não rejeitaríamos os rendimentos constantes de escala. Já que a distribuição do DEA-C/DEA-V possui freqüência elevada no índice 1.000.

No histograma abaixo observamos a distribuição da estatística  $d_i$ , observamos pelo histograma que quase não há freqüência para o índice diferente de 1, dizendo que as duas formas de medição, DEA-V e DEA-C são praticamente as mesmas<sup>71</sup>:

**Gráfico F.1. Histograma do  $d_i$  para a DEA-municípios.**



Fonte: Elaboração Própria.

<sup>71</sup> Em verdade, é preciso observar aqui que a concepção dessa segunda DEA, para municípios possui forte tendência para demonstrar rendimentos constantes. Observando-se os insumos, *rendapc*, *anal7a14*, *idh*, além da *pop2000*, *custo\_total* e número de *escolas*, percebemos uma forte correlação linear, que seria imprópria em uma estimação envolvendo erros aleatórios. Para o caso da DEA também não é muito elucidativa esse tipo de função de produção. Porém, testamos deixar apenas um de cada um dos insumos relacionados e a estimação ficou ainda assim parecida com a primeira. Optamos então por deixar todas as variáveis inicialmente pensadas.

Uma observação mais detalhada para observar quais são os municípios que não possuem  $d_i$  próximo a 1 demonstra que são sistematicamente os municípios grandes que se encontram nessa faixa de rejeição. Isso nos leva a crer que não rejeitamos a hipótese  $H_0$  de rendimentos constantes apenas porque esse segundo grupo é pequeno demais para viabilizar a rejeição de toda a distribuição, sendo mais provável que eles façam parte de um grupo distinto de municipalidades.

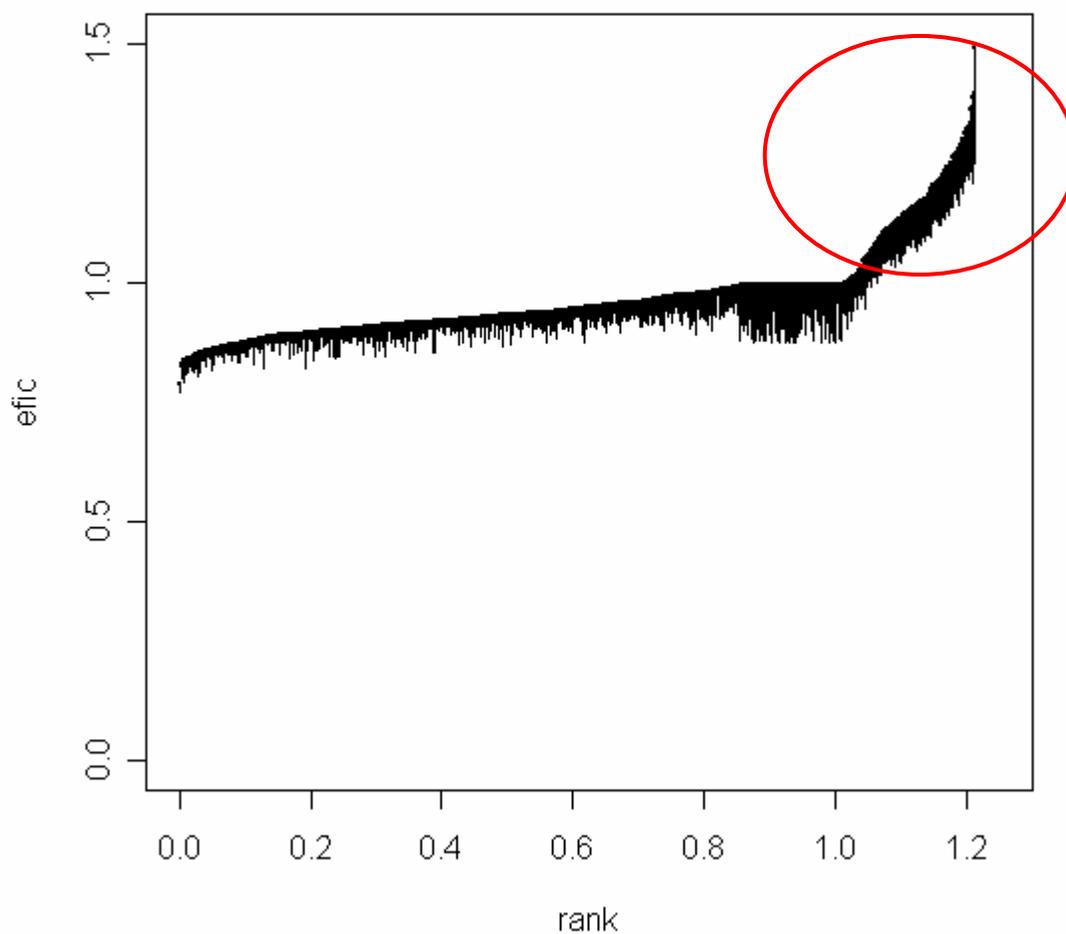
Na prática essa identificação estimando-se a DEA-C *bootstrap* com referência à uma fronteira variável de escala. Fornecemos no fim do anexo os comandos utilizados no R (É preciso baixar o pacote FEAR de Wilson 2005b, mais detalhes no fim desse anexo).

Há recurso visual onde podemos observar a característica diferente desses municípios rapidamente. O gráfico F.2, fornece os intervalos de confiança nas pequenas barras verticais demonstradas e o eixo da abscissa é a posição de acordo com o limite superior desse intervalo de confiança.

Observamos que os limites crescem num lentamente até alcançar os intervalos que estão em volta da eficiência  $\delta_i = 1$ . A partir dali, se nos deslocarmos um pouco mais para a direita nas posições do *rank*, observaremos uma ascensão vertiginosa dos intervalos, indicando um grupo diferente assinalado dentro da elipse vermelha.

Esse grupo é o que destoa da razão igual a 1 no histograma demonstrado à pouco e é o conjunto que se referencia à DEA-V e não à DEA-C. Em termos estatísticos o que ocorre é que a correção do viés nessa parte é negativo. Como a fórmula é  $\delta_i^* = \delta_i - \text{viés}$ . O índice de eficiência supera o limite superior igual a 1.

## Minas Gerais Eficiência por Municípios



Fonte: Elaboração Própria.

O ranking final de municípios foi construído de acordo com a elaboração acima e o resultados para os 100 primeiros estão nas páginas a seguir:

Colocação	Código	muníc	pop2000	analf_7_14a	estudo_11_25a	rendapc2000	pobreza	idh
1º	312870	GUAXUPE	47,036	8.62	8.44	233.84	25.45	0.739
2º	314480	NOVA LIMA	64,387	10.76	7.82	244.92	23.59	0.744
3º	313750	LAG DA FORMOSA	16,293	6.02	2.30	147.32	41.69	0.689
4º	310620	BELO HORIZONTE	2,238,526	10.43	15.79	414.94	18.89	0.791
5º	313760	LAG DA SANTA	37,872	15.76	6.42	206.49	39.03	0.728
6º	317020	UBERLANDIA	501,214	7.20	9.96	306.29	14.13	0.778
7º	316250	SÃO JOÃO DEL REI	78,616	10.59	7.48	189.04	39.66	0.727
8º	314310	MONTE CARMELO	43,899	10.32	4.19	207.24	32.08	0.717
9º	314930	PEDRO LEOPOLDO	53,957	16.68	7.42	197.24	35.46	0.748
10º	314710	PARA DE MINAS	73,007	16.22	5.57	202.18	33.99	0.728
11º	312710	FRUTAL	46,566	12.73	4.49	233.27	27.24	0.731
12º	310400	ARAXÁ	78,997	8.82	6.47	224.20	26.76	0.736
13º	313240	ITAJUBA	84,135	10.34	11.82	246.68	28.66	0.768
14º	313670	JUIZ DE FORA	456,796	10.23	11.58	311.64	21.08	0.769
15º	315250	POUSO ALEGRE	106,776	9.43	11.81	276.92	17.29	0.764
16º	313380	ITAUNA	76,862	9.37	7.59	206.52	25.34	0.737
17º	311330	CARANGOLA	31,921	17.05	7.11	153.37	45.19	0.697
18º	314590	OURO BRANCO	30,383	7.14	10.14	249.82	26.88	0.744
19º	312390	ENTRE RIOS DE MINAS	13,114	17.11	2.51	100.08	61.59	0.665
20º	311530	CATAGUASES	63,980	11.61	5.87	174.63	41.66	0.728
21º	317070	VARGINHA	108,998	7.86	13.07	295.26	23.05	0.772
22º	310420	ARCOS	32,687	8.66	3.78	169.16	37.71	0.717
23º	312510	EXTREMA	19,219	9.17	4.36	219.21	24.74	0.701
24º	315180	POCOS DE CALDAS	135,627	6.30	9.41	295.24	14.53	0.778
25º	312230	DMNOPO LIS	183,962	12.68	7.19	216.08	26.86	0.755
26º	316295	SÃO JOSE DALAPA	15,000	12.59	2.32	154.52	33.02	0.702
27º	315150	PIUMHI	28,783	8.02	4.60	205.28	26.10	0.710
28º	313420	ITUIUTABA	89,091	8.60	7.51	237.54	26.13	0.747
29º	316990	UBA	85,065	13.88	5.81	171.20	39.54	0.715
30º	313460	JABOTICATUBAS	13,530	31.87	1.59	117.24	58.20	0.631
31º	311120	CAMPO BELO	49,187	9.71	4.11	151.02	46.01	0.690
32º	316930	TRES CORACOES	65,291	15.10	9.18	210.02	33.06	0.717
33º	315960	SANTA RITADO SAPUCAI	31,264	16.47	9.05	219.06	39.51	0.716
34º	316370	SÃO LOURENÇO	36,927	8.58	6.41	208.84	38.29	0.743
35º	314800	PATOS DE MINAS	123,881	6.85	5.73	225.42	32.54	0.731
36º	316870	TIMOTÉO	71,478	10.28	7.05	221.46	28.70	0.768
37º	313820	LAVRAS	78,772	11.31	11.21	228.08	27.88	0.743
38º	310350	ARAGUARI	101,974	9.30	6.19	224.80	24.79	0.754
39º	313840	LEOPOLDINA	50,097	14.86	4.54	155.10	48.47	0.695
40º	310740	BOM DESPACHO	39,943	12.92	4.85	196.56	28.96	0.721
41º	315120	PIRAPORA	50,300	18.60	4.01	167.54	50.56	0.711
42º	314470	NOVA ERA	17,754	27.33	3.15	189.47	37.42	0.694
43º	311860	CONTAGEM	538,017	11.38	3.87	208.93	24.18	0.730
44º	316720	SETE LAGOAS	184,871	12.97	6.54	196.10	35.20	0.739
45º	317010	UBERABA	252,051	7.39	11.00	279.32	19.20	0.763

Colocação	Código	munic	pop2000	analf_7_14a	estudo_11_25a	rendapc2000	pobreza	ich
55ª	313170	ITABIRA	98,322	16,50	4,47	206,20	36,31	0,727
56ª	311110	CAMPINA VERDE	19,100	12,88	3,94	170,04	36,43	0,712
57ª	314560	OLIVEIRA	37,250	13,91	4,45	147,92	46,76	0,701
58ª	313130	IPATINGA	212,496	10,87	6,41	215,21	30,22	0,735
59ª	311430	CARMO DO PARANAIBA	29,460	10,83	3,66	172,21	34,68	0,709
60ª	312090	CURVELO	67,512	18,18	4,04	153,05	49,59	0,685
61ª	316040	SANTO ANTONIO DO MONT	23,473	10,49	5,07	178,23	37,75	0,698
62ª	316935	TRES MARIAS	23,568	13,39	1,66	149,32	38,51	0,708
63ª	314000	MARIANA	46,710	17,78	5,33	169,21	46,20	0,708
64ª	312410	ESMERALDAS	47,090	18,43	1,34	116,85	56,33	0,653
65ª	313940	MANHUACU	67,123	18,27	4,23	162,40	44,70	0,669
66ª	315780	SANTA LUZIA	184,903	13,15	2,04	146,70	36,69	0,686
67ª	311940	CORONEL FABRICIANO	97,451	13,58	5,27	191,17	35,06	0,720
68ª	312910	GURINHATA	6,883	14,86	1,55	137,96	46,87	0,681
69ª	310800	BOM SUCESSO	17,064	13,40	1,84	116,21	55,34	0,665
70ª	313440	ITURAMA	28,814	11,38	6,01	219,16	26,15	0,714
71ª	317130	MCOSA	64,854	13,12	13,10	224,79	37,38	0,723
72ª	315140	PITANGUI	22,269	10,83	4,34	158,39	37,86	0,709
73ª	314790	PASSOS	97,211	9,00	9,38	245,77	24,09	0,721
74ª	312160	DIAMANTINA	44,259	25,52	6,73	136,86	52,31	0,674
75ª	311420	CARMO DO CAJURU	17,157	10,49	2,63	124,96	41,78	0,690
76ª	317120	VESPASIANO	76,422	18,54	2,24	134,70	42,06	0,677
77ª	314390	MURIAE	92,101	13,07	5,71	185,38	43,93	0,707
78ª	313720	LAGOA DA PRATA	38,758	9,40	4,42	147,16	39,06	0,679
79ª	316500	SÃO TIAGO	10,245	14,60	2,25	104,31	63,53	0,659
80ª	316070	SANTOS DUMONT	46,789	16,20	4,05	152,36	46,87	0,697
81ª	317100	VAZANTE	18,928	11,12	3,35	157,14	40,40	0,686
82ª	310730	BOCAIÚVA	42,806	20,68	2,82	117,01	64,98	0,651
83ª	314110	MATOZINHOS	30,164	19,38	3,60	135,01	44,25	0,703
84ª	311160	CAMPOS GERAIS	26,541	16,07	3,95	138,51	47,55	0,662
85ª	314810	PATROCÍNIO	73,130	12,43	6,57	184,22	37,92	0,719
86ª	310540	BARÃO DE COCAIS	23,391	21,14	1,43	144,06	37,81	0,689
87ª	316940	TRES PONTAS	51,024	13,31	6,40	176,70	38,11	0,696
88ª	311650	CLARO DOS POÇOS	8,193	27,47	0,84	67,78	77,89	0,591
89ª	315460	RIBEIRAO DAS NEVES	246,846	18,07	1,14	122,19	43,87	0,674
90ª	312980	IBIRITÉ	133,044	17,82	0,81	115,78	45,47	0,668
91ª	314070	MATEUS LEITE	24,144	18,22	4,45	148,71	44,18	0,672
92ª	310560	BARBACENA	114,126	12,80	8,34	193,40	37,62	0,716
93ª	310670	BETIM	306,675	15,52	3,28	161,67	39,10	0,695
94ª	311340	CARATINGA	77,789	21,41	4,86	139,36	52,58	0,659
95ª	316970	TURMALINA	15,655	25,14	1,28	83,18	62,62	0,599
96ª	313510	JANAUBA	61,651	24,52	3,68	109,41	64,72	0,641
97ª	315210	PONTE NOVA	55,303	14,89	7,01	174,59	45,60	0,688
98ª	310150	ALÉM PARAIBA	33,610	15,79	8,30	202,09	32,48	0,715
99ª	316860	TEÓFILO OTONI	129,424	26,34	5,80	152,26	55,19	0,651
100ª	316840	TARUMIRIM	14,488	16,55	2,01	69,54	75,22	0,604

Por fim, o comando utilizado:

```
-----  
outm3 <- t(matrix(c(modelo3$pmate, modelo3$pport, modelo3$mat_total),nrow=809,ncol=3))  
inpm3 <- t(matrix(c(modelo3$pop2000, modelo3$Alfabet, modelo3$rendapc2000, modelo3$Sidh,  
modelo3$somaqdoc, modelo3$salee, modelo3$escolas),nrow=809,ncol=7))  
DEAM3<- dea(XOBS=inpm3,YOBS=outm3,ORIENTATION=2,RTS=3)  
BDEAM3<-  
boot.sw98(XOBS=inpm3,YOBS=outm3,DHAT=DEAM3,RTS=3,ORIENTATION=2,NREP=1000)  
  
DEAM_v3<- dea(XOBS=inpm3,YOBS=outm3,ORIENTATION=2,RTS=1)  
BDEAM_v3<-  
boot.sw98(XOBS=inpm3,YOBS=outm3,DHAT=DEAM_v3,RTS=3,ORIENTATION=2,NREP=1000)  
  
#A parte destacada em vermelho é indica que a referência de intervalo é o DEA-V, enquanto que a estimação  
é no DEA-C (RTS=3)  
-----
```

Observações finais:

1. A maioria dos procedimentos escritos nesse trabalho envolvem edições de comandos como este que estão relacionados acima. É comum o autor disponibilizar estas informações para quem se interesse em fazer trabalhos que dêem seqüência à metodologia que foi empregada. Todos os detalhes dessa dissertação podem ser requeridos.
2. Nessa dissertação, desenvolvemos os comandos para o “*jackstrap*” de Sampaio Stosic (2005). Em desenvolvimentos futuros tal lista de comandos poderá vir a se transformar em um pacote do software. Também se encontra disponível.
3. O R é um programa estatístico livre na linguagem GUI – *Graphical User Interface*, onde cada autor pode implementar suas rotinas no processo, como as que foram implementadas acima. Os autores do projeto pedem para ajudar na difusão do programa através dos dados:

R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>

4. O pacote estatístico principal utilizado para este trabalho foi o FEAR 0.913. de Wilson 2005b:

Paul W. wilson (2005). FEAR: Frontier Efficiency Analysis with R. R package version 0.913.