

# INFRA-ESTRUTURA E AVANÇOS EDUCACIONAIS

Marcelo Neri<sup>1</sup>  
Rodrigo Moura<sup>2</sup>  
Paulo Correa<sup>3</sup>

## 1. OBJETIVO

Este trabalho tem como objetivo dar um primeiro passo no sentido de avaliar impactos educacionais de mudanças no acesso à infra-estrutura a partir das bases de microdados oriundas da PNAD/IBGE e do SAEB/MEC. A dificuldade central destas análises correlações brutas e as parciais entre as variáveis de infraestrutura e as de desempenho escolar para fins de desenho de políticas é a determinação da direção de causalidade entre variáveis exógenas e as endógenas. Discutimos ao fim extensões em torno de análises de diferença em diferença deste impactos tomando como base a ocorrência de experiências onde algumas comunidades são beneficiadas com aumento da oferta de infraestrutura e outras não.

O presente projeto propõe desenvolver sistema de provisão de informação interativos e amigáveis, tais como simuladores de probabilidades desenvolvidos a partir de modelos logísticos (e.g., frequência escolar), além de simuladores de regressões de variáveis em nível (e.g., proficiência escolar) e banco de dados disponibilizados na internet.

Os indicadores sociais, portanto, a serem avaliados são: matrícula, frequência (assiduidade), progresso escolar (repetência) e proficiência. As variáveis explicativas a serem consideradas são de infra-estrutura, tais como:

1. Acesso a serviços públicos: abastecimento de água, esgotamento sanitário e iluminação elétrica e comunicações.
2. Infra-estrutura privada: número de pessoas no domicílio, direito de propriedade.

## 2. METODOLOGIA

Os principais elementos deste relatório consistem na avaliação das correlações entre acesso à infra-estrutura pública e privada conforme indicadores detalhados anteriormente e outros indicadores de desempenho escolar tais como matrícula, frequência, repetência e proficiência escolar a partir de técnicas micro-econômicas.

A metodologia da pesquisa consiste na geração, descrição e análise de um conjunto de base de dados. Utilizar-se-á de pesquisa domiciliar tradicional como a PNAD, que nos permite uma avaliação detalhada das correlações entre as condições de acesso à infra-estrutura doméstica em algumas variáveis sociais. Os efeitos sociais da provisão de infra-estrutura também será captado através de informações do Ministério da Educação por meio de avaliação de desempenho do SAEB. Nesse caso, estimar-se-á a equação de notas dos alunos, enfatizando os efeitos de variáveis ligadas ao acesso à infra-estrutura. Os resultados das regressões serão apresentadas em formato que permita a identificação imediata do impacto de cada variável. Todos os modelos empíricos gerados serão convertidos em ferramentas interativas e amigáveis a fim de aumentar a capacidade de difusão dos resultados da pesquisa.

---

<sup>1</sup> Centro de Políticas Sociais do IBRE e da EPGE.

<sup>2</sup> Doutorando da EPGE

<sup>3</sup> Banco Mundial

## **2.1. Técnicas**

### **Análise Bivariada**

O objetivo da análise bivariada é traçar um perfil da estrutura de correlações entre as variáveis de infra-estrutura e as de desempenho escolar, analisando o papel de cada atributo tomado isoladamente. Isto é, desconsideramos possíveis inter-relações das variáveis de desempenho com outras "variáveis explicativas."

### **Análise Multivariada**

A análise multivariada visa proporcionar um experimento melhor controlado que a análise bivariada. Seu objetivo é captar o padrão de correlações parciais entre as variáveis de interesse e as variáveis explicativas. Na análise multivariada captamos as correlações das variáveis de acesso à infra-estrutura com variáveis de retorno social citadas mantendo as demais variáveis constantes.

Trabalharemos com duas variantes do modelo de regressão multivariada:

- a) Regressão em mínimos quadráticos ordinários para variáveis contínuas;
- b) Regressão logística binomial multinomial envolvendo como endógenas duas ou mais categorias de variáveis discretas.

## **2.2. Bases de Dados**

### **Sistema Nacional de Avaliação da Educação Básica (SAEB)**

O Sistema Nacional de Avaliação da Educação Básica (SAEB) avalia os alunos brasileiros da 4a. e da 8a. série do ensino fundamental e da 3a. série do ensino médio em Língua Portuguesa e Matemática. Além das provas das duas disciplinas, questionários coletaram informações sobre as características da escola, do diretor, do professor, da turma e dos alunos. Infra-estrutura está incluída em variáveis de background familiar.

Especificamente, estudaremos nessa base de dados, o desempenho escolar médio dos alunos da 4ª e 8ª série do ensino fundamental e 3ª série do ensino médio medido através de proficiência e de atraso escolar função das seguintes variáveis: tem eletricidade, tem água encanada.

### **Pesquisa Nacional por Amostra de Domicílios (PNAD)**

A pesquisa abrange todo o Brasil, exceto as áreas rurais de Acre, Amapá, Amazonas, Rondônia, Roraima e Pará. Contém informações sobre diversas características demográficas e sócio-econômicas da população. Especificamente, utilizaremos as seguintes variáveis:

- i. **Educacionais:** sabe ler e escrever, frequenta escola ou creche, série e espécie de curso que frequenta;
- ii. **Serviços públicos:** água proveniente de rede geral de distribuição, esgotamento sanitário (rede geral), iluminação elétrica.

## **2.3 Modelo de análise**

O desempenho dos alunos pode ser afetado pelo acesso a serviços públicos como luz, água etc. Melhor infra-estrutura, tanto na casa como na escola, deve melhorar a produtividade dos estudantes, reduzindo portanto, a repetência. Mas a questão que se levanta aqui é que outros fatores podem afetar estas variáveis de interesse. Assim, os

controles assumidos ajudam a isolar de forma mais precisa o efeito da infra-estrutura no desempenho escolar. Assim, por exemplo, a educação dos pais pode afetar positivamente a produtividade de seus filhos, mesmo que haja certa deficiência em infra-estrutura na escola. Além disso, o número de moradores na casa do aluno pode também influenciar. Temos observado que há geralmente um número ótimo de moradores que ajudam a melhorar a performance escolar do estudante. Além disso, controlamos também por sexo, cor e Unidade Federativa (UF), visto que as características escolares podem divergir entre esses grupos.

Assim, os modelos a serem estimados se baseiam na equação geral:

$$y_i = F(x\beta)$$

em que,  $y_i$  é a variável dependente, que são os indicadores de desempenho escolar, para os quais será estimada regressões separadamente;  $x$  é o vetor dos controles, tal que  $x=(IND; FAM; ESC; INF; ENV)$ . IND é o sub-vetor que contém as características dos indivíduos, FAM as características da família, ESC da escola, INF as variáveis de infra-estrutura referente à moradia, e ENV inclui alguma externalidade (por exemplo, favela). A seguir, discutimos brevemente o modelo para cada um dos indicadores, apontando para potenciais e problemas e suas possíveis soluções.

### *Proficiência*

As regressões para a equação de proficiência foram estimadas por série. O modelo estimado foi um modelo linear nas co-variáveis. Estimamos a regressão por mínimos quadrados ordinários, mas consideramos para a hipótese de heterocedasticidade dos erros. Assim, estimamos a matriz de variância através do estimador de White (1980) produzindo erros padrões robustos a heterocedasticidade.

### *Reprovação e assiduidade*

Para as variáveis independentes reprovação e assiduidade utilizamos um modelo logit multinomial, sendo que o valor base de comparação é nunca ter reprovado e nunca ter faltado à escola. O modelo estimado por máxima verossimilhança é o que segue:

$$P(y = j | x) = \frac{\exp(x\beta_j)}{1 + \sum_{h=1}^J \exp(x\beta_h)}, j = 1, 2, J = 2$$

em que,  $y$  é a variável reprovação, o vetor  $\beta_j$  é o conjunto de parâmetros para  $j=1$  (reprovou uma vez) e  $j=2$  (reprovou 2 ou mais vezes). Como as probabilidades devem somar um, temos que ter:

$$P(y = 0 | x) = 1 / \left[ 1 + \sum_{h=1}^J \exp(x\beta_h) \right], j = 1, 2, J = 2$$

Assim, sempre comparamos em relação à base ( $y=0$ ) que é nunca reprovou. Podemos simplificar a notação da probabilidade de resposta como:

$$p_j(x, \beta) = P(y = j | x)$$

$$p_0(x, \beta) = P(y = 0 | x)$$

Deve-se ressaltar que a interpretação da magnitude dos parâmetros estimados deste modelo não é direta. O efeito marginal decorrente de uma mudança em uma variável controle contínua é:

$$\begin{aligned}
\frac{\partial p_j(x, \beta)}{\partial x_k} &= \beta_{jk} \frac{\exp(x\beta_j)(1 + \sum_{h=1}^J \exp(x\beta_h)) - \exp(x\beta_j)(1 + \sum_{h=1}^J \beta_{hk} \exp(x\beta_h))}{(1 + \sum_{h=1}^J \exp(x\beta_h))^2} \\
&= \frac{\exp(x\beta_j)}{(1 + \sum_{h=1}^J \exp(x\beta_h))} \left\{ \beta_{jk} - \frac{(1 + \sum_{h=1}^J \beta_{hk} \exp(x\beta_h))}{(1 + \sum_{h=1}^J \exp(x\beta_h))^2} \right\} \\
&= p_j(x, \beta) \left\{ \beta_{jk} - \frac{(1 + \sum_{h=1}^J \beta_{hk} \exp(x\beta_h))}{(1 + \sum_{h=1}^J \exp(x\beta_h))^2} \right\}
\end{aligned}$$

Além disso, através da razão das probabilidades em relação à base temos:

$$\frac{p_j(x, \beta)}{p_0(x, \beta)} = \exp(x\beta_j), j = 1, 2, J = 2$$

ou ainda:

$$\log[p_j(x, \beta) / p_0(x, \beta)] = x\beta_j$$

Ou seja, temos uma interpretação mais direta de uma variação de uma unidade em  $x$ , que mostra o quanto varia o log da razão das probabilidades (*log-odds*), através do parâmetro estimado.

### Matrícula e atraso

Para a variável matrícula e atraso, estimamos um logit por máxima verossimilhança. O modelo estimado aqui é um caso particular do logit multinomial para quando  $J=2$ , ou seja:

$$\Pr(y = 1 | x) = \frac{\exp(x\beta)}{1 + \exp(x\beta)}$$

em que,  $y$  é a matrícula,  $x$  é o vetor dos controles e  $\beta$  o vetor dos parâmetros. O efeito parcial deste modelo para uma mudança em uma variável contínua é:

$$\frac{\partial \Pr(y = 1 | x)}{\partial x_j} = \frac{\exp(x\beta)}{(1 + \exp(x\beta))^2} \beta_j = \frac{\Pr(y = 1 | x)}{1 + \exp(x\beta)} \beta_j$$

Assim, o sinal do efeito parcial é dado pelo sinal do parâmetro. Se  $x_k$  for discreta, o efeito parcial é obtido mudando  $x_k$  de um valor para outro, *ceteris paribus*, e calculando a diferença da probabilidade. No caso de  $x$  ser variável binária teríamos:

$$\frac{\exp(\beta_1 + \beta_2 x_2 + \dots + \beta_k x_k)}{1 + \exp(\beta_1 + \beta_2 x_2 + \dots + \beta_k x_k)} - \frac{\exp(\beta_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1})}{1 + \exp(\beta_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1})}$$

E conhecendo o sinal de  $\beta_k$ , é suficiente para determinar o sinal do efeito parcial. Mas deve-se ressaltar que para encontrar a magnitude do efeito dever-se-ia estimar as equações acima.

### 2.4 Apresentação dos dados

Nesta seção apresentamos os dados que irão alimentar o modelo e também na análise bivariada. Os dados foram retirados da Pesquisa Nacional por Amostra e Domicílios (PNAD) e do Sistema Nacional de Avaliação da Educação Básica (SAEB). Em relação ao SAEB, os dados estão agrupados para alunos, professores, diretores e escola. Utilizamos informações das bases dos alunos e da escola, e unimos estes dados no nível do aluno. A PNAD divide sua base para pessoas e domicílios. As variáveis de infra-

estrutura são extraídas da parte de domicílios. Assim, juntamos os dados no nível das pessoas.

Para a proficiência e reprovação, as variáveis independentes utilizadas, respectivamente, foram, segundo os grupos definidos:

- IND: sexo (0 = homem e 1= mulher); dummies para o atributo cor<sup>4</sup> (d\_cor1=branca d\_cor2=parda d\_cor3=negra d\_cor4=amarela d\_cor5=indígena) e idade.
- FAM: educmae (1=nunca estudou 2=<4serie 3=4s 4=<8s 5=8s 6=<EnsinoMedio 7=EM 8=começou mas não terminou faculdade 9=faculdade 10=não sei); educpai (mesma descrição); reg\_metr(0=mora em região metropolitana e 1=não); dummies para as unidades federativas<sup>5</sup> (UF)
- ESC: comp (número total de computadores disponíveis para os alunos na escola); conshidra (conservação das instalações hidráulicas na escola :1=adequado 2=regular 3=inadequado 4=inexistente); conseletr (conservação das instalações elétricas na escola :1=adequado 2=regular 3=inadequado 4=inexistente); ilumina (1= as salas de aula são iluminadas e 0=não); local (local da escola 0=urbana e 1=rural); rede (0=particular 1=pública).
- INF: cpunet(1= tem computador com internet em casa e 0=não tem), cpu (1= tem computador em casa e 0= não tem); nres (número de pessoas residentes no domicílio do aluno) ; luz/água (onde você mora existe eletricidade/água encanada? 0= não e 1=sim)

Em relação as variáveis matrícula, atraso e assiduidade, foram utilizados dados da Pnad. Na análise bivariada e multivariada, trabalhamos com as seguintes variáveis<sup>6</sup>, segundo os grupos definidos:

- IND: sexo (0 = homem e 1= mulher); dummies para o atributo cor<sup>7</sup> (branca, parda e negra) e idade.
- FAM: moramae (a mãe mora junto? 0=não, 1=sim); local (local do domicílio 0=urbana e 1=rural); setor (setor do domicílio 0= não especial 1= aglomerado subnormal (favela)).
- INF: nres (número de pessoas residentes no domicílio); nres\_10 (número de pessoas residentes no domicílio com 10 anos ou mais); cpunet(1= tem acesso à internet em casa e 0=não tem), cpu (1= tem computador em casa e 0= não tem); água (água para pelo menos 1 cômodo? 0= não e 1=sim); dummies para a forma de recebimento da luz (d\_luz1= elétrica, d\_luz2=óleo, querosene ou gás de botijão, d\_luz3=outra forma); telefone(0= não tem telefone fixo e 1= tem); celular (0= não tem e 1= tem), dummies para a condição do domicílio<sup>8</sup> (proppago=próprio já pago e terreno próprio, proppagando= próprio pagando e terreno próprio, alugado= domicílio alugado, cedido=domicílio cedio por empregador ou de outra forma, e outra condição).

Vale ressaltar que a variável assiduidade somente estava disponível na Pnad de 2001.. Além disso, a Pnad não contém dados sobre características da escola, por isso o subvetor ESC é nulo.

---

<sup>4</sup> Nas regressões, o atributo omitido foi a cor indígena.

<sup>5</sup> Nas regressões, a UF omitida foi Rondônia.

<sup>6</sup> As variáveis celular, cpu e cpunet não têm na PNAD de 1999.

<sup>7</sup> Nas regressões, o atributo omitido foi a cor indígena.

<sup>8</sup> A dummy omitida aqui foi a d\_condicao1 (próprio, já pago).

### 3. RESULTADOS

#### 3.1 Análise Bivariada da Proficiência e da Reprovação

Nesta seção apresentamos a relação das variáveis de infra-estrutura tomadas isoladamente nas variáveis de interesse. A proficiência é a nota do aluno no exame de Matemática. Esta disciplina consegue medir de uma forma mais adequada a produtividade (ou performance) do aluno em sua escola. Outra variável de interesse abordada é a reprovação. Vale destacar que, na análise da variável reprovação (referente à questão “Você já foi reprovado?”), esta apresenta os seguintes valores:

- 0 = nenhuma reprovação
- 1 = uma reprovação
- 2 = duas ou mais reprovações

A seção segue por grupos de variáveis de infra-estrutura: comunicações (acesso à internet, computador etc), serviços públicos (acesso à luz, água etc), proficiência versus reprovação e outros fatores. Para cada grupo fizemos a análise separada para proficiência e reprovação, com exceção da análise conjunta entre essas variáveis.

##### 3.1.1 Comunicações

###### *Proficiência*

As conseqüências do acesso a computadores em casa são relativamente desconhecidas, e particularmente, a literatura dos impactos educacionais de computadores em casa é escassa.

A importância de se analisar o acesso de computadores em casa reside no fato de que existe evidência na literatura (Selwyn, 1998) que seu uso em casa pode compensar a falta da tecnologia da informação na escola. Assim, na tabela abaixo, observa-se uma nota (proficiência) maior para os alunos que tem acesso à Internet em casa, via computador, para todos os anos, tendo um diferencial maior em relação aos que não tem, no ano mais recente. Discriminado por série este impacto (ver Apêndice) nota-se um diferencial crescente, entre os que tem e não tem, em relação a série. Ou seja, a diferença da nota entre quem tem acesso e não tem é maior para quem está na 3ª série do Ensino Médio, do quem está na 8ª série do Ensino Fundamental, que é maior em relação a quem está na 4ª série do Ensino Fundamental. Mesmo considerando os que tem computador em casa, sem discriminar se tem acesso à internet ou não, o desempenho foi melhor para aqueles que possuem.

2001

Você tem em sua casa acesso à internet? -----	média(profic) -----	N(profic) -----
Sim	250.8451	1607576
Não	220.2743	6692015

2003

na sua casa tem computador com internet? -----	média(profic) -----	N(profic) -----
Sim	272.4335	1540403
Não	215.7434	7367194
não sei	177.5196	101,979

Da mesma forma, avaliamos o acesso a computadores pelos alunos na escola. Observe da tabela abaixo, que, em geral, quanto maior o número de computadores disponíveis para os alunos melhor o desempenho dos mesmos na escola.

1999

computadores para uso dos alunos -----	média(profic) -----	N(profic) -----
1 a 5	237.602	56,644.90
6 a 10	227.2147	31,339.30
11 a 15	225.6821	98,483.30
16 a 20	247.5213	6,671.29
21 a 30	237.8901	1455028
mais de 30	217.7308	6734760

2001

computadores para uso dos alunos -----	média(profic) -----	N(profic) -----
1 a 5	205.0509	4680556
6 a 10	232.8716	693,674
11 a 15	243.7516	1466733
16 a 20	246.2613	395,596
21 a 30	277.228	386,129
mais de 30	237.3867	1135969

2003

computadores para uso dos alunos -----	média(profic) -----	N(profic) -----
1 a 5	227.3424	1365810
6 a 10	233.5717	1487797
11 a 15	249.8801	828,026
16 a 20	244.1725	513,874
21 a 30	256.5461	466,069
mais de 30	302.1497	387,577

Uma estatística exclusiva para o ano de 2001 que inferimos é se o aluno usa o computador para fazer a lição de casa. O que notou-se é que não há uma relação muito direta entre o uso do computador para fazer a lição e a proficiência dos alunos. Apenas no caso do aluno nunca utilizar, este têm uma nota, em média, menor em relação aos que usam, principalmente para a 8ª série e a 3ª série.

### *Reprovação*

Na tabela abaixo, corroborando o resultado apresentado para a proficiência, observa-se uma maior repetência para quem não tem computador com acesso à internet. Em todas séries, (ver Apêndice) observa-se o mesmo resultado para todas as séries.

2001

Você tem em sua casa acesso à internet? -----	média(reprova) -----	N(reprova) -----
Sim	0.40198	1546560
Não	0.5710047	6366278

2003

na sua casa tem computador com internet? -----	média(reprova) -----	N(reprova) -----
sim	0.2768215	1528659
não	0.5112	7286987
não sei	0.6421474	99,591.70

Bem como para análise na escola, o maior acesso a computadores reduz a média de repetência entre os alunos, principalmente em 2003.

1999

computadores para uso dos alunos -----	média(reprova) -----	N(reprova) -----
1 a 5	0.5474932	56,327.20
6 a 10	0.4352457	31,339.30
11 a 15	1.047137	98,169.50
16 a 20	0.3861432	6,548.49
21 a 30	0.7390051	1429809
mais de 30	0.5980511	6561561

2001

computadores para uso dos alunos -----	média(reprova) -----	N(reprova) -----
1 a 5	0.6081356	4343471
6 a 10	0.5498762	668,522
11 a 15	0.5121428	1421098
16 a 20	0.4229966	375,891
21 a 30	0.3581446	374,782
mais de 30	0.4677808	1068423

2003

computadores para uso dos alunos -----	média(reprova) -----	N(reprova) -----
1 a 5	0.4646084	1350097
6 a 10	0.4088988	1474182
11 a 15	0.4253787	819,876
16 a 20	0.3830172	509,141
21 a 30	0.3528308	462,670
mais de 30	0.2554964	384,601

Em relação ao uso de computadores na realização de tarefas escolares, evidenciou-se que: (i) para alunos da 4ª série há uma relação direta “paradoxal” em relação à reprovação, ou seja, quanto mais se usa maior o índice médio de reprovação; (ii) para a 8ª série e 3ª série os maiores índices são referentes àqueles que nunca e sempre usam. Assim pode-se concluir que o uso do computador para fazer tarefas escolares pode reduzir a reprovação e até melhorar o desempenho do aluno, mas desde que o mesmo não se utilize deste artifício sempre.



## Serviços Públicos

### *Proficiência*

O acesso à infra-estrutura não tem sido objeto de estudo explícito na literatura. Apenas tem servido como controle para se avaliar o efeito de algum fator em alguma variável educacional (performance, matrícula etc). Então torna-se imprescindível focar a análise em aspectos de acesso à infra-estrutura, pois o mesmo afeta o desempenho escolar. A questão que surge (e será dirimida na seção da análise multivariada) é o quanto que o acesso a luz, água etc e melhora a proficiência escolar, tendo-se controlado para outros fatores (como educação dos pais, unidade federativa, número de residentes na mesma casa etc). Observa-se das tabelas que o uso da eletricidade e da água tem uma correlação positiva com a proficiência e há um peso mais negativo sobre quem não tem acesso à eletricidade do que à água.

1999

onde você mora existe eletricidade? -----	média(profic) -----	N(profic) -----
Não	167.8943	222,185
Sim	223.2689	8062880
onde você mora chega água pela torneira? -----	média(profic) -----	N(profic) -----
Não	189.4924	738,207
Sim	224.8105	7562802

2001

onde você mora existe eletricidade? -----	média(profic) -----	N(profic) -----
Sim	225.5403	8383812
Não	184.7264	116,958
onde você mora chega água pela torneira? -----	média(profic) -----	N(profic) -----
Sim	228.5357	7697211
Não	188.1576	808,658

2003

onde você mora existe eletricidade? -----	média(profic) -----	N(profic) -----
Sim	226.7333	8740838
Não	168.6937	295,599
onde você mora chega água pela torneira? -----	média(profic) -----	N(profic) -----
Sim	227.595	8426927
Não	186.714	601,508

Analizamos também algumas variáveis de infra-estrutura ligadas à escola. Da tabela abaixo observamos que a conservação das instalações hidráulicas e elétricas tem uma correlação positiva com o desempenho dos alunos<sup>9</sup>.

2001

---

<sup>9</sup> Apenas nas condições de regular e inadequado há uma certa inconsistência nos dados.

estado de conservação das instalações hidráulicas	média(profic)	N(profic)
Adequado	230.1152	5190654
Regular	214.5678	2375805
Inadequado	213.5871	1031459
Inexistente	148.339	81,439.90
estado de conservação das instalações elétricas	média(profic)	N(profic)
Adequado	228.2398	5412978
Regular	213.111	2148441
Inadequado	219.0134	1093401
Inexistente	169.0744	28,345.10

### 2003

estado de conservação das instalações hidráulicas	média(profic)	N(profic)
Adequado	231.9369	4928297
Regular	211.8014	2477907
Inadequado	223.6232	1455046
Inexistente	177.029	77,278.70
estado de conservação das instalações elétricas	média(profic)	N(profic)
Adequado	231.014	5440354
Regular	208.9853	2221690
Inadequado	223.8385	1287688
Inexistente	200.47	1,099.86

Outro aspecto importante está ligado ao fato das iluminação no local de estudo, como por exemplo, as salas de aula. As tabelas abaixo corroboram para este fato, mostrando um melhor desempenho nas salas iluminadas.

### 1999

as salas de aula são iluminadas?	média(profic)	N(profic)
Não	200.3453	750,353
Sim	223.6014	7506782

### 2001

as salas de aula são iluminadas?	média(profic)	N(profic)
Sim	225.1335	7927559
Não	199.9669	757,295

### 2003

as salas de aula são iluminadas?	média(profic)	N(profic)
Sim	226.4951	8157987
Não	203.3845	776,155

## Reprovação

O acesso à infra-estrutura também reduz o índice de reprovação, mas ao contrário do desempenho escolar, a eletricidade não gera um impacto mais negativo do que o acesso à água, em 2001 e 2003.

### 1999

onde você mora existe eletricidade? -----	média(reprova) -----	N(reprova) -----
Não	0.9134332	213,420
Sim	0.6180525	7889863
onde você mora chega água pela torneira? -----	média(reprova) -----	N(reprova) -----
Não	0.698581	721,023
Sim	0.6204154	7400020

### 2001

onde você mora existe eletricidade? -----	média(reprova) -----	N(reprova) -----
Sim	0.5423108	7945756
Não	0.7374656	107,850
onde você mora chega água pela torneira? -----	média(reprova) -----	N(reprova) -----
Sim	0.5397894	7304947
Não	0.6089589	747,563

### 2003

onde você mora existe eletricidade? -----	média(reprova) -----	N(reprova) -----
Sim	0.4657926	8650882
Não	0.6700777	290,648
onde você mora chega água pela torneira? -----	média(reprova) -----	N(reprova) -----
Sim	0.4577966	8339352
Não	0.6719981	594,071

Apesar da falta de robustez em relação à infra-estrutura escolar, geralmente colégios com instalações mais adequadas apresentam menor índice de repetência.

### 2001

estado de conservação das instalações hidráulicas -----	média(reprova) -----	N(reprova) -----
Adequado	0.4800252	4927499
Regular	0.6512823	2221137
Inadequado	0.6577147	961,470
Inexistente	0.8602442	68,605.30
estado de conservação das instalações elétricas -----	média(reprova) -----	N(reprova) -----
Adequado	0.4863783	5129538
Regular	0.6773081	2000685
Inadequado	0.6226653	1025419

Inexistente	0.555383	25,344.80
-------------	----------	-----------

### 2003

estado de conservação das instalaÇões hidráulicas	média(reprova)	N(reprova)
-----	-----	-----
Adequado	0.4454069	4877840
Regular	0.5228759	2446824
Inadequado	0.4858542	1442037
Inexistente	0.4907354	75,977

  

estado de conservação das instalaÇões   elétricas	média(reprova)	N(reprova)
-----	-----	-----
Adequado	0.4436818	5384944
Regular	0.5232141	2194337
Inadequado	0.5142236	1273841
Inexistente	0.5854016	1,080

Nota-se também que salas de aula iluminadas produzem uma menor média de repetência.

### 1999

as salas de aula são iluminadas?	média(reprova)	N(reprova)
-----	-----	-----
Não	0.7409022	720,548
Sim	0.6147693	7340966

### 2001

as salas de aula são iluminadas?	média(reprova)	N(reprova)
-----	-----	-----
Sim	0.5404339	7487220
Não	0.663017	694,647

### 2003

as salas de aula são iluminadas?	média(reprova)	N(reprova)
-----	-----	-----
Sim	0.4665973	8069782
Não	0.5494646	768,317

### Proficiência x Reprovação

Nesta subseção cruzamos os dados das duas variáveis que serão analisadas afim de adquirir mais *insight* para análise. Claramente quem nunca reprovou tem um desempenho muito melhor em relação a quem já reprovou. Mas em 2003, existe, a priori, um resultado não robusto, visto que a nota média de que repetiu duas vezes ou mais é maior em relação a quem repetiu mais de uma vez. Mas isso é devido a uma quantidade grande de alunos na 4ª série do EF que repetiram uma vez, série na qual apresenta-se as notas menores, puxando portanto a média da proficiência para baixo. Assim, no Apêndice, segue as notas por série para uma análise mais robusta. E nela observa-se que sempre quem repete mais obtém uma nota menor.

1999

Você já repetiu de ano? Quantas vezes?	média(profic)	N(profic)
0	232.8534	4660424
1	210.4588	1916071
2 ou +	204.6553	1607259

2001

Você já repetiu de ano? Quantas vezes?	média(profic)	N(profic)
Não	235.6762	5098051
sim, uma vez	213.4472	1778427
sim, duas vezes ou mais	205.6537	1375708

2003

you já foi reprovado?	média(profic)	N(profic)
Não	234.8306	5842136
sim, uma vez	204.1426	2008789
sim, duas vezes ou mais	210.6046	1119564

Outros fatores

### *Proficiência*

Além dos aspectos analisados que podem afetar o desempenho escolar, repetência, matrícula e frequência podemos incorporar variáveis regionais, como por exemplo, se a escola situa-se em área urbana ou rural. Este fator é importante visto que muitos programas sociais parecem ter um efeito bem maior na área rural, sendo que a mesma tem evoluído no Brasil, pelo menos em termos de frequência escolar.

Da tabela abaixo se observa um diferencial significativo em termos de produtividade escolar entre as escolas no meio rural e urbano, evidenciando claramente uma qualidade bem melhor da escola urbana em relação à rural, muito mais devido à deficiência desta última. Um dos elementos candidatos a explicar este diferencial de desempenho seria a diferença de caesso a infraestrutura entre áreas rural e urbana. Observa-se uma pequena evolução de 2001 para 2003 nas notas dos dois tipos de escola, sendo 1.83% da rural e 1.2% da urbana.

1999

Local	média(profic)	N(profic)
Urbana	223.419	8110314
Rural	164.9225	272,613

2001

Local	média(profic)	N(profic)
Urbana	225.9887	8410273
Rural	150.1539	348,383

2003

Local	média(profic)	N(profic)
Urbana	228.6984	8586694
Rural	152.9064	482,323

Outra deficiência que se apresenta no país que citamos brevemente é a da escola pública em relação à particular, como é notado nos valores de proficiência observado abaixo. As duas redes de ensino evoluíram pouco de 2001 para 2003 (em torno de 0.8%).

1999

rede	média(profic)	N(profic)
pública	213.8598	7233261
particular	269.6909	1149666

2001

rede	média(profic)	N(profic)
pública	214.2539	7664984
particular	284.0748	1093672

2003

rede	média(profic)	N(profic)
pública	215.9496	7946949
particular	286.4118	1122068

Analisamos também o fato do início dos estudos dos alunos. Esta questão está apenas relacionada para o ano de 2003. As tabelas abaixo estão por série e não foram agregadas, visto que a resposta “maternal” estava disponível apenas para os alunos da 4ª série. Observa-se a forte influência de se começar a estudar no maternal ou na pré-escola. Conforme Heckman (2005) as habilidades, tanto cognitivas como não cognitivas, das crianças são formadas nos primeiros anos de vida, se tornando menos maleáveis ao decorrer dos anos (principalmente as cognitivas), elevando, portanto, a produtividade das crianças. Portanto, a educação precoce influencia fortemente nesta formação de habilidades.

-> serie = 4

Quando você começou a estudar?	média(profic)	N(profic)
No maternal	193.7027	978,107
Na Pré-escola	178.5029	1833527
Na 1a série	159.9368	807,205
Na 2a série	140.5765	42,864.60
Na 3a série	131.3782	39,379.20

-> serie = 8

Quando você começou a estudar?	média(profic)	N(profic)
--------------------------------	---------------	-----------

-----	-----	-----
Na Pré-escola	250.2382	2509924
Na 1a série	226.3967	619,627
Na 2a série	224.0938	29,040.50
Na 3a série	201.2923	12,557.60

-> serie = 11

-----	-----	-----
Quando você começou a estudar?	média(profic)	N(profic)
-----	-----	-----
Na Pré-escola	284.0487	1646444
Na 1a série	260.46	423,162
Na 2a série	260.1382	19,859.50
Na 3a série	257.2694	7,693.49

### *Reprovação*

O índice de reprovação para escolas urbanas é bem menor do que as rurais, principalmente para as rurais. Vale ressaltar que, em 2003, existem apenas escolas rurais da 4ª série do EF.

1999

local	média(reprova)	N(reprova)
-----	-----	-----
urbana	0.6191831	7918775
rural	0.8582471	264,979

2001

local	média(reprova)	N(reprova)
-----	-----	-----
urbana	0.5404642	7954840
rural	0.775314	297,346

2003

local	média(reprova)	N(reprova)
-----	-----	-----
urbana	0.4649568	8490186
rural	0.6253288	480,303

O índice de reprovação para todas as séries é bem maior nas escolas públicas do que privadas, apesar de que, de 2001 para 2003 as reprovações caíram bastante na rede pública de ensino.

1999

rede	média(reprova)	N(reprova)
-----	-----	-----
pública	0.6818536	7045915
particular	0.2867776	1137839

2001

rede	média(reprova)	N(reprova)
------	----------------	------------

	média(reprova)	N(reprova)
pública	0.6004056	7181768
particular	0.2035369	1070418

2003

rede	média(reprova)	N(reprova)
pública	0.5136786	7855772
particular	0.1906985	1114717

A mensagem da tabela abaixo diz que quanto mais cedo as crianças começarem a estudar menor será o índice de repetência. O único resultado ambíguo apresentado é que quando se pergunta para alunos da 8ª série do EF e da 3ª série do EM, o índice de reprovação maior para os que começaram na 2ª série do EF e não na 3ª série.

-> serie = 4

Quando você começou a estudar?	média(reprova)	N(reprova)
No maternal	0.2830364	970,206
Na Pré-escola	0.3943322	1821329
Na 1a série	0.5875055	800,552
Na 2a série	0.6996055	41,956.50
Na 3a série	0.7794674	39,017.10

-> serie = 8

Quando você começou a estudar?	média(reprova)	N(reprova)
Na Pré-escola	0.4546385	2494947
Na 1a série	0.6585677	615,979
Na 2a série	1.019318	28,795.90
Na 3a série	0.9180655	12,309.80

-> serie = 11

Quando você começou a estudar?	média(reprova)	N(reprova)
Na Pré-escola	0.4827031	1640420
Na 1a série	0.6712016	421,498
Na 2a série	1.006604	18,919.40
Na 3a série	0.7552114	7,541.98

### 3.2 Análise Bivariada de Matrícula, Atraso e Assiduidade

A análise bivariada estende-se agora à matrícula, atraso e assiduidade. As variáveis de infra-estrutura encontradas na PNAD diferem um pouco das encontradas no SAEB, vistas na análise precedente. Aqui, restringimos a população de interesse à faixa etária de 5 à 17 anos que estão na “idade escolar ativa”. Não consideramos todas as pessoas, pois



estimativas iniciais se mostraram desconformes ou pouco robustas devido ao viés amostral<sup>10</sup>.

Os valores para estas variáveis são: matrícula e atraso (0= não e 1=sim), e assiduidade (0=nunca faltou à escola, 1=faltou pouco – 1 a 5 dias, 2= faltou muito – 6 ou mais dias).

## Serviços Públicos

### *Matrícula*

Em relação aos serviços públicos, notamos que o acesso à água gera maiores índices de matrícula. E a luz recebida na forma elétrica também incentiva à frequência escolar. Além disso, no contexto dinâmico, a média de matrícula escolar tem crescido entre aqueles que tem acesso à água e luz elétrica, ao longo de todos anos.

#### 1999

água para pelo menos 1 comodo?	média(frequenta)	N(frequenta)
Não	0.8176481	4959570
Sim	0.8898703	22000000
luz	média(frequenta)	N(frequenta)
elétrica	0.8820219	25200000
óleo, querosene ou gás de botijão	0.8047988	1682280
outra forma	0.701266	90522

#### 2001

água para pelo menos 1 comodo?	média(frequenta)	N(frequenta)
Não	0.8337093	7790659
Sim	0.9115559	35100000
luz	média(frequenta)	N(frequenta)
elétrica	0.903088	40400000
óleo, querosene ou gás de botijão	0.8082067	2287583
outra forma	0.7571673	182286

#### 2003

água	média(frequenta)	N(frequenta)
Não	0.842093	6682578
Sim	0.920597	36400000
luz	média(frequenta)	N(frequenta)
elétrica	0.912764	41200000
óleo, querosene ou gás de botijão	0.819523	1702814

<sup>10</sup> Nas análises bivariadas iniciais observou-se diferença praticamente nula da frequência média entre aqueles que, por exemplo, tinham acesso a água em relação aos que não tinham; ou ainda uma média maior para aqueles que não tinham este acesso. Este viés provavelmente foi causado devido à incorporação muito maior de pessoas que tem acesso à água e não estão mais frequentando a escola, ou seja, fora da idade escolar ativa.

outra forma	0.715452	144717
2004		
agua para pelo menos 1 comodo?	média(frequenta)	N(frequenta)
-----	-----	-----
Não	0.8392105	7103324
Sim	0.9243643	37700000
-----	-----	-----
luz	média(frequenta)	N(frequenta)
-----	-----	-----
elétrica	0.9166356	42700000
óleo, querosene ou gás de botijão	0.798355	1921788
outra forma	0.7322503	168721

### *Atraso*

Para a variável atraso, temos a mesma dinâmica que a apresentada para a matrícula, em favorecimento a quem tem acesso à água e luz elétrica. E tem também decaído ao longo dos anos.

1999		
agua para pelo menos 1 comodo?	média(atraso)	N(atraso)
-----	-----	-----
Não	0.6899563	4056799
Sim	0.5221236	19600000
-----	-----	-----
luz	média(atraso)	N(atraso)
-----	-----	-----
elétrica	0.5402294	22300000
óleo, querosene ou gás de botijão	0.7220165	1353897
outra forma	0.6352237	63480

2001		
agua para pelo menos 1 comodo?	média(atraso)	N(atraso)
-----	-----	-----
Não	0.6733003	6495717
Sim	0.4810638	32000000
-----	-----	-----
luz	média(atraso)	N(atraso)
-----	-----	-----
elétrica	0.5023136	36500000
óleo, querosene ou gás de botijão	0.7267892	1848840
outra forma	0.6219706	138021

2003		
agua	média(atraso)	N(atraso)
-----	-----	-----
Não	0.6367887	5628259
Sim	0.4569119	33500000
-----	-----	-----
luz	média(atraso)	N(atraso)
-----	-----	-----

elétrica	0.4746866	37600000
óleo, querosene ou gás de botijão	0.6942151	1395713
outra forma	0.578348	103538

## 2004

água para pelo menos 1 comodo?	média(atraso)	N(atraso)
-----	-----	-----
Não	0.6189638	5961184
Sim	0.4451128	34900000
Luz	média(atraso)	N(atraso)
-----	-----	-----
Elétrica	0.46152	39200000
óleo, querosene ou gás de botijão	0.6914837	1534269
outra forma	0.5715523	123546

## Assiduidade

Relembrando que para assiduidade, um valor maior implica em maiores falta, ou seja, menor assiduidade, observamos um resultado abaixo contraditório àquelas discutidos mais acima, no sentido de que quem tem acesso a estes serviços públicos a frequência escolar é menor..

## 2001

água para pelo menos 1 comodo?	média(assiduidade)	N(assiduidade)
-----	-----	-----
Não	0.5020853	6494164
Sim	0.5403322	31900000
-----	-----	-----
Luz	média(assiduidade)	N(assiduidade)
-----	-----	-----
Elétrica	0.5345272	36400000
óleo, querosene ou gás de botijão	0.5079964	1848840
outra forma	0.7066968	138021

## Comunicações

### Matrícula

Continuando a análise das conseqüências do uso do computador na educação, verificamos, a priori, que o fato de possuir este bem em casa gera maiores incentivos à matrícula escolar, mas que a diferença em relação a quem não tem, vem decrescendo ao longo dos anos, mas permanece significativa. Em relação ao uso da internet, observamos o mesmo comportamento e trajetória.

## 2001

Tem computador em casa?	média(frequenta)	N(freqüenta)
-----	-----	-----
Não	0.887654	38300000
Sim	0.9804623	4497002
-----	-----	-----
Este computador tem acesso à internet?	média(frequenta)	N(freqüenta)

-----	-----	-----
Não	0.8909383	40000000
Sim	0.9856133	2882380

### 2003

Tem computador em casa?	média(frequenta)	N(frequenta)
-----	-----	-----
Não	0.89803	37600000
Sim	0.980128	5447262
-----	-----	-----
Este computador tem acesso à internet?	média(frequenta)	N(frequenta)
-----	-----	-----
Não	0.900644	39200000
Sim	0.986287	3868050

### 2004

Tem computador em casa?	média(frequenta)	N(frequenta)
-----	-----	-----
Não	0.9003832	38800000
Sim	0.9786938	6002659
-----	-----	-----
Este computador tem acesso à internet?	média(frequenta)	N(frequenta)
-----	-----	-----
Não	0.9032069	40700000
Sim	0.9827948	4261276

Na relação entre telefone e matrícula escolar, nota-se que os agentes que tem maior acesso a telefone fixo, em todos anos, bem como para o celular.

### 1999

telefone	média(frequenta)	N(frequenta)
-----	-----	-----
0	0.8795639	26400000
1	0.9558904	18400000

### 2001

telefone	média(frequenta)	N(frequenta)
-----	-----	-----
0	0.855336	23600000
1	0.9488261	19300000
celular	média(frequenta)	N(frequenta)
-----	-----	-----
1	0.9003277	41700000
0	0.7910335	1147136

### 2003

telefone	média(frequenta)	N(frequenta)
-----	-----	-----
0	0.873589	24200000
1	0.952903	18900000
-----	-----	-----

celular	média(frequenta)	N(frequenta)
0	0.891022	28300000
1	0.941929	14700000

## 2004

telefone	média(frequenta)	N(frequenta)
0	0.8795639	26400000
1	0.9558904	18400000

celular	média(frequenta)	N(frequenta)
0	0.886922	24800000
1	0.940552	20000000

## Atraso

Para a variável atraso, observamos que a média de atraso é maior entre aqueles que não tem computador em casa e nem acesso à internet.

## 2001

cpu	média(atraso)	N(atraso)
0	0.5276967	34000000
1	0.40422	4409740

cpunet	média(atraso)	N(atraso)
0	0.5219815	35700000
1	0.4072516	2841511

## 2003

Tem computador em casa?	média(atraso)	N(atraso)
Não	0.4958333	33800000
Sim	0.4003049	5339015

cpunet	média(atraso)	N(atraso)
0	0.4918897	35300000
1	0.3995494	3815008

## 2004

Tem computador em casa?	média(atraso)	N(atraso)
Não	0.4826318	35000000
Sim	0.3982484	5874765

Este computador tem acesso à internet?	média(atraso)	N(atraso)
Não	0.4790056	36700000
Sim	0.3946031	4187960

O mesmo comportamento observamos para aqueles que não tem telefone fixo e/ou celular.

1999

telefone	média(atraso)	N(atraso)
-----	-----	-----
0	0.5963552	16000000
1	0.4559377	7668578

2001

telefone	média(atraso)	N(atraso)
-----	-----	-----
0	0.5789664	20200000
1	0.4414183	18300000
celular	média(atraso)	N(atraso)
-----	-----	-----
1	0.5085987	37500000
0	0.7177566	907635

2003

telefone	média(atraso)	N(atraso)
-----	-----	-----
0	0.5327811	21100000
1	0.42425	18000000
celular	média(atraso)	N(atraso)
-----	-----	-----
0	0.5080833	25300000
1	0.4366927	13900000

2004

telefone	média(atraso)	N(atraso)
-----	-----	-----
0	0.5124188	23300000
1	0.4150193	17600000
celular	média(atraso)	N(atraso)
-----	-----	-----
0	0.5034124	22000000
1	0.4320223	18800000

### *Assiduidade*

Em relação à assiduidade, observamos que quem tem computador e internet, bem como telefonia fixa são mais assíduos. Apenas quem tem celular parece faltar à escola.

2001

cpu	média(assiduidade)	N(assiduidade)
-----	-----	-----
0	0.5450747	34000000
1	0.4474161	4406069
cpunet	média(assiduidade)	N(assiduidade)
-----	-----	-----

0	0.5428345	35600000
1	0.4258538	2840346

## 2001

telefone	média(assiduidade)	N(assiduidade)
0	0.548281	20100000
1	0.5179874	18300000
celular	média(assiduidade)	N(assiduidade)
0	0.5023479	907423
1	0.5346315	37500000

## Infra-estrutura privada

### *Matricula*

Em relação, primeiramente, ao número de cômodos nota-se um relação crescente com a frequência. Ou seja, famílias com maior *background* de infra-estrutura em sua casa, gera maiores incentivos à freqüentar escolas. Já em relação a variável moradores, observamos que existe uma quantidade ótima de moradores que seria em torno de quatro ou cinco moradores. Após isso, provavelmente, as crianças e adolescentes deixam de ir à escola, afim de trabalharem para contribuírem na renda de casa afim de expandir suas restrições orçamentárias.

## 1999

Numero de comodos no domicilio	média(frequenta)	N(frequenta)
1	0.7420627	209109
2	0.7766051	733732
3	0.8255257	2208692
4	0.8454088	4261058
5	0.8800182	7071854
6	0.8844594	5550128
7	0.901104	3096041
8	0.9250169	1735431
9	0.9246513	889505
10 ou mais	0.9383388	1245855
moradores	média(frequenta)	N(frequenta)
1	0.859843	530355
2	0.8190097	1205308
3	0.862588	3238524
4	0.9083726	7094408
5	0.8942238	5936724
6	0.8760253	3668047
7	0.8580535	2196025
8	0.8403407	1331679
9	0.8274537	746252
10 ou mais	0.8021911	1054083

## 2001

Numero de comodos no domicilio	média(frequenta)	N(frequenta)
1	0.7491831	292552
2	0.7964185	1196209
3	0.8357358	3410463
4	0.8662214	6571449
5	0.8984153	11000000
6	0.9115904	8569825
7	0.9197211	5209386
8	0.9349114	2939378
9	0.9443455	1524514
10	0.9620727	2249458
moradores	média(frequenta)	N(frequenta)
1	0.6523692	20703
2	0.8086602	1118450
3	0.8965989	5162548
4	0.9288501	11900000
5	0.9099749	10300000
6	0.8939955	5979800
7	0.8710197	3381795
8	0.8586481	2152883
9	0.8485966	1210349
10	0.8152452	1762098

### 2003

comodo	média(frequenta)	N(frequenta)
0	0.841241	47676
1	0.744	243762
2	0.812194	895104
3	0.8521	3076917
4	0.879945	6283230
5	0.906596	11400000
6	0.917944	8939524
7	0.930738	5297075
8	0.938857	3138058
9	0.946062	1557226
10	0.97316	2185717
nres	média(frequenta)	N(frequenta)
1	0.567391	17265
2	0.836081	1221426
3	0.912097	5737593
4	0.936871	12600000
5	0.924139	9803992
6	0.891362	5722288
7	0.877789	3332775
8	0.859879	1964053
9	0.880947	1113120
10	0.83317	1568211

### 2004

Numero de comodos no domicilio	média(frequenta)	N(frequenta)
--------------------------------	------------------	--------------



	média(frequenta)	N(frequenta)
1	0.7391282	275599
2	0.8246477	1028752
3	0.8641669	3166054
4	0.8846498	7035590
5	0.910362	12000000
6	0.9205015	9011824
7	0.9279163	5405774
8	0.9422122	3178438
9	0.9519637	1532985
10	0.9682235	2232684
moradores	média(frequenta)	N(frequenta)
1	0.6716926	24672
2	0.8521075	1274047
3	0.9099234	6057666
4	0.9368107	13200000
5	0.9268593	10300000
6	0.9006416	5847590
7	0.8866862	3256621
8	0.8697135	2037426
9	0.857598	1082176
10	0.8270499	1755084

Em relação à titularidade, observamos um resultado interessante: quem tem casa própria e ainda está pagando, apresenta maiores taxas de matrícula. E quem mora em terreno próprio segue nesta mesma direção. Ou seja, quem já tem seu terreno, mas ainda está pagando a moradia própria parece gerar maiores incentivos a seus filhos (5 a 17 anos) a irem à escola.

#### 1999

prop	média(frequenta)	N(frequenta)
1	0.8852199	17300000
2	0.8581356	1609051
3	0.9285918	1385472
4	0.9038534	153266
5	0.867522	3171266
6	0.8272851	3384416

#### 2001

prop	média(frequenta)	N(frequenta)
1	0.9089996	27500000
2	0.8633094	2721768
3	0.9444129	1898014
4	0.9134105	153633
5	0.8824748	5117237
6	0.8529015	5450261

#### 2003

prop	média(frequenta)	N(frequenta)
------	------------------	--------------

1	0.918274	28000000
2	0.88398	2349075
3	0.943581	1845245
4	0.91549	203148
5	0.893723	5391861
6	0.869656	5306483

#### 2004

prop	média(frequenta)	N(frequenta)
-----	-----	-----
1	0.9198748	29100000
2	0.8872339	2442444
3	0.9495703	1759597
4	0.918296	196955
5	0.9035383	5958344
6	0.8677395	5340431

#### *Atraso*

Para a variável atraso, nota-se uma relação decrescente entre o número de cômodos e o atraso, ou seja, quanto maior a infra-estrutura da casa, menor o índice de repetência. E para o número de moradores, observamos o mesmo comportamento que as taxas de matrícula, tendo um valor ótimo perto de 3 ou 4 moradores.

#### 1999

Numero de comodos no domicilio	média(atraso)	N(atraso)
-----	-----	-----
1	0.6227348	155172
2	0.5905918	569820
3	0.578396	1823894
4	0.5628356	3604557
5	0.5578312	6223749
6	0.5712692	4909673
7	0.543408	2791260
8	0.5184385	1605803
9	0.4742307	823015
10 ou mais	0.4361044	1171199
-----	-----	-----
moradores	média(atraso)	N(atraso)
-----	-----	-----
1	0.5963089	456022
2	0.6080293	988255
3	0.5164934	2794868
4	0.4659683	6446053
5	0.5216406	5309554
6	0.6023767	3215307
7	0.6598758	1885405
8	0.6686874	1119064
9	0.6859588	617489
10 ou mais	0.7106066	846125

#### 2001

Numero de comodos no domicilio	média(atraso)	N(atraso)
1	0.5648956	219175
2	0.5440521	953043
3	0.5162625	2850246
4	0.535893	5694906
5	0.5199415	9844665
6	0.5268003	7812170
7	0.5107299	4791748
8	0.4838089	2748058
9	0.4733567	1440267
10	0.4255622	2164142
moradores	média(atraso)	N(atraso)
1	0.5730046	13506
2	0.5514182	904446
3	0.458287	4629095
4	0.4356106	11000000
5	0.4980779	9343145
6	0.5631033	5347712
7	0.6144245	2945610
8	0.652915	1849135
9	0.6836563	1027098
10	0.6722355	1436754

## 2003

comodo	média(atraso)	N(atraso)
0	0.5654374	40107
1	0.5119184	181359
2	0.5245351	726998
3	0.4785532	2621842
4	0.4987277	5529380
5	0.4863301	10400000
6	0.4969668	8206789
7	0.477922	4930187
8	0.4740823	2946187
9	0.4312281	1473232
10	0.4170035	2127052
nres	média(atraso)	N(atraso)
1	0.9010821	9796
2	0.5456659	1021211
3	0.4175749	5233725
4	0.4159759	11800000
5	0.4739346	9060464
6	0.5399646	5100631
7	0.585992	2925692
8	0.6029721	1689448
9	0.6381359	980600
10	0.6343412	1306587

## 2004

Numero de comodos no domicilio	média(atraso)	N(atraso)
--------------------------------	---------------	-----------

1	0.5412242	203703
2	0.5088241	848358
3	0.4628386	2735999
4	0.491811	6224033
5	0.4690473	11000000
6	0.4840782	8295398
7	0.4747882	5016106
8	0.4472367	2994763
9	0.4309081	1459346
10	0.3989033	2161737
moradores	média(atraso)	N(atraso)
1	0.6056601	16572
2	0.5210621	1085625
3	0.4139354	5512012
4	0.4110319	12400000
5	0.4625213	9578557
6	0.513551	5266583
7	0.5549115	2887601
8	0.6036179	1771977
9	0.6148833	928072
10	0.6242383	1451542

Observamos para a titularidade do terreno do domicílio, o mesmo comportamento anterior. Mas com um detalhe a mais: quem já tem seu domicílio quitado e construído em terreno próprio apresenta um atraso maior em relação aos demais. Ou seja, os filhos apresentam um comportamento mais cômodo em relação a sua situação.

1999

prop	média(atraso)	N(atraso)
1	0.561671	15300000
2	0.5966313	1380784
3	0.4669941	1287663
4	0.5162853	138530
5	0.5121284	2751677
6	0.548158	2799877

2001

prop	média(atraso)	N(atraso)
1	0.5255452	25000000
2	0.5620731	2350294
3	0.3854608	1792509
4	0.462802	140330
5	0.4649982	4515833
6	0.5226291	4650546

2003

prop	média(atraso)	N(atraso)
1	0.498165	25700000
2	0.5396721	2076535

3	0.3626804	1741139
4	0.3581353	185980
5	0.4365447	4819037
6	0.4702275	4614817

2004

prop	média(atraso)	N(atraso)
-----	-----	-----
1	0.4860532	2680000
2	0.496411	2167019
3	0.3828523	1670861
4	0.3467818	180863
5	0.4198767	5383592
6	0.463526	4634103

### *Assiduidade*

Para assiduidade, novamente maior quantidade de cômodos leva a maior frequência. Já o número de moradores não apresenta exatamente a mesma relação anterior. Mas famílias numerosas apresentam uma assiduidade elevada.

2001

Numero de comodoss no domicilio	média(assiduidade)	N(assiduidade)
-----	-----	-----
1	0.578784	219175
2	0.6085991	952683
3	0.5993659	2849054
4	0.566748	5691316
5	0.5406826	9833923
6	0.5268086	7809627
7	0.5221883	4789330
8	0.4988635	2746511
9	0.4550192	1438537
10	0.4458025	2164142
moradores	média(assiduidade)	N(assiduidade)
-----	-----	-----
1	0.1849548	13506
2	0.613258	903486
3	0.5648911	4625350
4	0.5288798	1100000
5	0.5389259	9338072
6	0.5346892	5343383
7	0.5272954	2943553
8	0.4910912	1848569
9	0.4903841	1025851
10	0.4965716	1436542

Já a variável assiduidade apresenta uma relação oposta à variável atraso, visto que quem mora em moradias quitadas e terreno próprio apresenta maior assiduidade.

2001

prop	média(assiduidade)	N(assiduidade)
------	--------------------	----------------

1	0.5090867	25000000
2	0.5563452	2348896
3	0.556349	1791665
4	0.5240861	140330
5	0.5985276	4510955
6	0.5844589	4648121

Outros fatores

### *Matrícula*

A variável “local” é específica para o local onde os indivíduos moram. E corrobora-se os esperado, o qual, pessoas que moram em centros urbanos tem um maior índice de matrícula em relação aos que moram na zona rural.

Em relação à variável “setor”, que diz se a residência do morador está situada em um setor não especial ou em um aglomerado subnormal (favela). Nota-se que, mesmo em condições habitacionais adversas, os indivíduos tendem a buscar mais educação, não muito abaixo em relação a quem mora em setores não especiais (normais).

1999

local	média(frequenta)	N(frequenta)
0	0.8892497	21000000
1	0.8323189	5997361
setor	média(frequenta)	N(frequenta)
0	0.8769016	25800000
1	0.8701541	1188232

2001

local	média(frequenta)	N(frequenta)
0	0.9096634	34900000
1	0.8435276	8004544
setor	média(frequenta)	N(frequenta)
0	0.8984092	41000000
1	0.8744507	1934316

2003

local	média(frequenta)	N(frequenta)
0	0.918899	35300000
1	0.860545	7800171
setor	média(frequenta)	N(frequenta)
0	0.90895	41200000
1	0.895352	1936356

2004

local	média(frequenta)	N(frequenta)
-------	------------------	--------------

0	0.923244	36000000
1	0.8600178	8867280
setor	média(frequenta)	N(frequenta)
0	0.9110292	42900000
1	0.9051089	2031243

### *Atraso*

Nota-se abaixo que quem mora na zona rural apresenta altos índices de atraso escolar em relação aos habitantes urbanos. Também quem mora em favelas apresenta dificuldades na escola e conseqüentemente gerando maiores atrasos. Mas nota-se para todos os casos que o atraso escolar tem decaído ao longo dos anos.

#### 1999

local	média(atraso)	N(atraso)
0	0.5284899	18700000
1	0.6346558	4993333
setor	média(atraso)	N(atraso)
0	0.5492237	22600000
1	0.5871019	1034507

#### 2001

local	média(atraso)	N(atraso)
0	0.4898743	31800000
1	0.6246523	6752832
setor	média(atraso)	N(atraso)
0	0.5119646	36800000
1	0.546982	1692030

#### 2003

local	média(atraso)	N(atraso)
0	0.4631112	32400000
1	0.5784101	6713097
Setor	média(atraso)	N(atraso)
0	0.4819192	37400000
1	0.5035926	1733721

#### 2004

local	média(atraso)	N(atraso)
0	0.449045	33300000
1	0.5634206	7626019
setor	média(atraso)	N(atraso)

0	0.469429	39100000
1	0.4903176	1838496

*Assiduidade*

Para assiduidade, observamos que quem mora na zona rural é um pouco mais assíduo. Mas quando olha-se para os habitantes de favela, estes apresentam menor frequência escolar.

2001

local	média(assiduidade)	N(assiduidade)
-----	-----	-----
0	0.5377131	31700000
1	0.5173526	6750508
setor	média(assiduidade)	N(assiduidade)
-----	-----	-----
0	0.5330999	36800000
1	0.5568419	1690505



## **Análise Multivariada**

### *Proficiência*

Das tabelas<sup>11</sup> abaixo das regressões por mínimos quadrados ordinários de variáveis contínuas, observa-se que a falta de eletricidade tem um efeito mais perverso que a água, na grande maioria dos casos. O impacto da infra-estrutura escolar (conservação das instalações hidráulicas e elétricas e iluminação na sala de aula) também afetam positivamente a proficiência escolar. O acesso a computadores pode desenvolver no aluno uma melhor “aptidão escolar”, gerando bons resultados na escola. O acesso à internet em casa não apresenta uma direção clara, mas parece que em alguns casos pode gerar uma queda da produtividade do aluno. Quando comparamos o impacto de se ter computador em casa com o impacto de se ter computador na escola disponível para o aluno, o primeiro fator tem peso, em termos de magnitude, bem maior que o segundo. Isso provavelmente é causado pelo melhor ambiente encontrado em casa frente à escola, e pelo fato do aluno ter uma menor competição pelo uso em casa em relação à escola. Logo, em termos gerais, a infra-estrutura pode impactar positivamente na proficiência escolar, podendo até chegar a um aumento em torno de 12 pontos na nota (no caso de se ter computador em casa, em relação a quem não tem, *ceteris paribus*). Em relação à infra-estrutura privada, observamos o comportamento esperado: um efeito positivo em nível, e efeito negativo no termo quadrático.

---

<sup>11</sup> A dummy omitida para cor foi a indígena, e a omitida para a UF foi Rondônia. Não nos preocupamos em omitir determinada dummy, visto que os parâmetros de interesse são os relacionados ao acesso à serviços públicos (infra-estrutura).

	1999				2000				
	sem infra	infra	infra + interações	sem infra	infra	infra + interações	sem infra	infra	infra + interações
comp	0.1869665	0.300485	0.2423366	3.554587	2.946311	3.020156	-1.50372	-1.70219	-1.748065
ilumina	3.002468	2.333415	2.323744	4.14988	3.803128	3.875708	7.193788	6.404406	6.476974
local	-0.6104815	0.998699	-10.09922	(dropped)	(dropped)	(dropped)	(dropped)	(dropped)	(dropped)
rede	-29.82373	-25.6331	-12.35598	-27.8376	-23.76565	7.108875	-24.3598	-22.1381	-20.75181
cpu		10.0562	-11.78205		11.45336	-31.03031		8.935715	-21.49608
luz		7.908338	22.10887		14.4223	41.99128		19.07808	36.48184
agua		1.064662	-3.724729		1.991611	-6.648823		5.669532	-11.88166
calcamento		1.091363	1.042587		-0.3006738	-0.3389106		1.318064	1.443282
mres		5.435737	10.91749		6.551465	12.62957		7.629351	16.50101
mres2		-1.17096	-1.236257		-1.252153	-1.289062		-1.47484	-1.483715
luzagua			11.29179			19.73966			16.66497
luzcpu			9.837572			26.99686			21.41842
aguacpu			12.98823			15.91528			9.258737
luzmres			-3.083532			-5.26546			-6.702328
aguanres			-2.630618			-0.7477434			-2.29744
luzeducmae			-0.2080364			-2.463471			-15.53619
aguaducpai			1.908606			1.009844			4.105757
luzeducpai			3.064654			0.4055832			9.140925
aguareducpai			3.134212			1.338479			3.826199
redelocal			9.646005			(dropped)			(dropped)
luzlocal			0.5920121			(dropped)			(dropped)
luzrede			-9.200353			-17.82483			0.8337453
aguadlocal			-0.2882066			(dropped)			(dropped)
aguarede			-3.40846			-13.51657			-2.060484

	2001								
	sem infra	infra	infra + interações	sem infra	Infra	infra + interações	sem infra	infra	infra + interações
comp	1.229546	1.218326	1.148187	1.931808	1.75348	1.719981	2.139192	1.886654	1.871466
conshidra	-0.4907568	0.058317	-0.3008023	-0.34104	-0.326323	-0.2843197	-0.36572	-0.41052	-0.3661442
conseletr	-1.611983	-1.77326	-1.508786	-2.60737	-2.232286	-2.176497	-4.12652	-3.83565	-3.785436
ilumina	0.4017318	0.767631	0.7661395	0.793191	0.7497473	0.8926448	2.631798	1.99109	2.134213
local	-6.815586	-6.68513	-40.57292	-5.25604	-8.047334	33.5506	(dropped)	(dropped)	(dropped)
rede	-32.28158	-28.4072	-22.23629	-31.8333	-27.59194	-13.09592	-31.8917	-28.6223	-5.848006
cpunet		-6.26504	-6.163741		0.4417026	-18.66986		2.396598	-4.149301
cpu		11.7414	3.035438		11.2534	-14.07341		12.75787	19.5711
luz		-2.38887	(dropped)		13.59094	3.876835		11.64352	18.363
agua		7.593086	13.76029		4.024484	3.48233		3.755718	11.08742
calcamento		(dropped)	5.412716		-0.9046958	-0.4922245		-1.38105	-1.088251
mres		8.638517	11.8512		2.767683	2.707576		1.910306	6.813336
mres2		-1.37443	-1.408189		-0.5260364	-0.5489251		-0.39707	-0.4060748
luzagua			-4.271273			4.103571			1.600342
luzcpu			(dropped)			21.76282			-6.893004
aguacpu			9.583749			3.701233			-0.0963633
luzcpunet			(dropped)			8.40684			-9.505528
aguacpunet			-0.250999			10.81229			16.04394
luzmres			-2.103145			1.288532			-4.196873
aguamres			-1.042282			-1.136921			-0.7388201
luzeducmae			1.641162			2.963528			1.287357
agueducmae			1.456901			2.640709			0.9325508
luzeducpai			6.169796			0.3574734			7.764943
agueducpai			1.378469			2.444105			5.286523
redelocal			36.93553			(dropped)			(dropped)
luzlocal			2.830498			-76.94223			(dropped)
luzrede			-1.451547			-4.79093			-3.118315
aguaindustrial			-9.044202			36.60086			(dropped)
aguarede			-5.171471			-9.794577			-20.28205

	2003								
	sem infra	infra	infra + interações	sem infra	infra	infra + interações	sem infra	infra	infra + interações
comp	2.308094	2.011055	2.101788	3.113396	2.858523	2.88872	2.953441	2.660939	2.832032
conshidra	1.368222	1.718548	1.919746	-2.35751	13.29951	-2.530864	-2.52367	-2.10338	-1.998266
consetr	-4.816982	-4.65178	-4.753675	-0.2333	-2.478793	-0.0239627	-4.47958	-4.74984	-4.769175
ilumina	5.562978	5.307833	5.04849	3.177513	-0.0755579	2.271246	2.033465	0.897933	1.037007
local	-2.768282	-1.38279	-6.15396	(dropped)	2.325165	(dropped)	(dropped)	(dropped)	(dropped)
rede	-27.50066	-24.6541	-19.62867	-26.8254	(dropped)	-23.98865	-27.7233	-24.1228	-14.29632
reg_metr	-0.6524825	-0.18315		-2.08079	-23.22444		-6.62813	-5.71628	
cpunet		1.768082	5.342111		-1.392861	23.61476		10.57108	-14.32877
cpu		6.589169	-8.999372		9.626012	-26.25557		3.395497	-9.852057
luz		16.29624	-6.715517		3.074709	-8.761472		13.31873	33.85883
agua		4.722835	18.76682		0.3190562	-19.90762		5.383378	-7.976632
mres		-3.68909	-1.444951		3.704906	4.343506		2.151946	6.944927
mres2		0.073908	0.0627739		-0.4726964	-0.4540884		-0.37082	-0.3700736
luzagua			10.24127			23.61476			14.66839
luzcpu			15.0445			-26.25557			7.256142
aguacpu			0.7268683			-8.761472			6.317746
luzcpunet			-7.605997			-19.90762			-0.55661
aguacpunet			3.868659			4.343506			25.96557
luzmres			0.833995			-0.4540884			-1.806461
aguanres			-3.036654			0.4160586			-3.13936
luzeducmae			0.3499199			3.337236			-5.639199
agueducmae			0.2168198			-0.7226317			3.414945
luzeducpai			1.445262			-1.026148			2.278738
agueducpai			-0.2874098			1.264063			-2.111788
redereg_metr			3.696077			-1.314251			-4.371282
redelocal			0.9699498			(dropped)			(dropped)
luzlocal			-0.869903			(dropped)			(dropped)
luzrede			1.195668			8.323944			-13.6682
aguarede			3.927584			(dropped)			(dropped)
aguarede			-8.147804			-6.788772			6.670517

## *Reprovação*

A título de exemplificação, interpretando a variável *nres* no ano de 1999, temos que, um morador a mais no domicílio aumenta o log da razão por 0.141289 (comparando reprovado uma vez com nenhuma vez), ou seja, reduz-se a probabilidade de não ser reprovado e/ou aumenta-se a probabilidade de ser reprovado uma vez. Das variáveis de infra-estrutura, observamos que a maioria é estatisticamente não significativa. Somente *cpu*, *cpunet*, e *água*, *nres* e *nres2* em alguns casos são significativas. Destas somente o acesso a computador e internet apresentam o sinal negativo e de acordo com o esperado (com exceção de 1999), ou seja, quem tem computador e internet em casa, tem uma maior chance de não ser reprovado e uma menor chance de ser reprovado. As interações na grande maioria não são estatisticamente válidas e sua inclusão reduz o efeito das outras e/ou tornando-as também não significativas, principalmente na 8ª série e na 3ª série.

	1999				2000				
	sem infra	infra	infra + interações	sem infra	infra	infra + interações	sem infra	infra	infra + interações
comp	-0.0296071	-0.04962	-0.0485481	-0.00326	0.0215616	0.0188594	0.015699	0.030312	0.0343769
ilumina	-0.1002844	-0.07168	-0.0700756	-0.10412	-0.1367382	-0.1363008	0.02943	0.039493	0.0298805
local	-0.0491383	-0.09615	-0.0047466						
rede	0.8311396	0.684054	1.122512	0.473997	0.3849079	0.0839433	0.293653	0.250856	2.34367
cpu		-0.33971	1.538532		-0.2799723	0.4614338		-0.08835	-0.0103078
luz		0.082741	0.5456143		0.0475325	-0.5364261		-0.14524	0.9833294
agua		0.024528	-1.115977		0.0703138	0.123358		0.097116	-1.573797
calcamento		-0.18342	-0.1840836		0.0553542	0.0573113		0.052362	0.0466504
mes		0.141289	-0.0832483		-0.2141298	-0.5453859		-0.10978	-0.3650873
mes2		0.00377	0.0063088		0.0453617	0.0461324		0.037348	0.0379893
luzagua			0.281431			-0.4411053			1.515237
luzcpu			-1.559823			-0.3966524			-0.6222508
aguacpu			-0.3990523			-0.3458557			0.5623138
luzmes			0.0433734			0.2609706			0.1179848
aguanmes			0.2033872			0.074129			0.1381778
luzeducmae			-0.2081621			-0.067623			0.0690168
agueducmae			-0.0160651			0.0249831			-0.268835
luzeducpai			0.1162215			-0.1357045			-0.05505
agueducpai			-0.0665405			-0.1538653			-0.214471
redelocal			-0.1243515						
luzlocal			0.0463632						
luzrede			-0.6505395			-0.0116414			-2.490329
agualocal			0.0360254						
aguarede			0.1979048			0.3168735			0.3951434

	1999					
	sem infra		infra		sem infra	
	Infra	infra + interações	8ª série EF	infra + interações	3ª série EM	infra + interações
2						
comp	0.0199859	0.0220842	-0.07262	-0.0854969	-0.02454	0.0472225
ilumina	-0.0837143	-0.066318	0.102753	0.0567091	0.022335	0.0344306
local	-0.0575828	0.1584995				
rede	1.115088	1.799451	0.577249	0.7600801	0.364029	2.081342
cpu	-0.20892	1.783288	-0.256725	1.594128	-0.48098	0.2606725
luz	0.026671	-0.2307872	0.2774347	-1.700873	-0.42889	-0.5088818
agua	0.027521	-0.3270885	0.3578211	-0.3126254	0.221991	-0.2260378
calçamento	-0.0041	0.0052268	0.0767304	0.0736837	0.223703	0.2166924
mres	0.014213	-0.3020278	-0.0368409	-0.7420088	-0.19639	-0.9906028
mres2	0.024072	0.0179071	0.0290874	0.0303574	0.053741	0.052811
luzagua		-0.0217763		0.1302857		-0.4351379
luzcpu		-1.515299		-1.244247		-0.5701526
aguacpu		-0.5964338		-0.6359795		-0.1892721
luzmres		0.1207309		0.5559148		0.5039302
aguanmres		0.328023		0.1636631		0.3177665
luzeducmae		-0.1625942		0.2862969		1.30041
aguadeducmae		-0.2093651		-0.0046631		-0.0050858
luzeducpai		-0.1384181		-0.1821797		-0.7103169
aguadeducpai		0.0725218		-0.0781634		-0.3604715
redelocal		-0.4292656				
luzlocal		0.0557075				
luzrede		0.0487606		-0.3761006		-2.123307
agualocal		0.2382103				
aguarede		-0.9120324		0.0892985		0.2999401

	2001								
	sem infra	infra	infra + interações	sem infra	infra	infra + interações			
1		4ª série EF		8ª série EF		3ª série EM			
comp	-0.026442	-0.03398	-0.0328384	-0.03113	-0.0245374	-0.0231678	-0.02208	-0.01812	-0.0183658
conshidra	-0.0048876	-0.0109	-0.0089876	0.024691	0.0313861	0.0293151	-0.00615	0.00924	0.009447
conseletr	0.038101	0.046174	0.0419757	0.031418	0.0292671	0.0287585	0.057699	0.048327	0.0473894
ilumina	-0.0907745	-0.06407	-0.0616707	0.18359	0.2131436	0.2087108	-0.03039	-0.01985	-0.0222095
local	0.0271869	0.052554	0.3667054	-0.29859	-0.1931014	23.87385			
rede	0.7406878	0.632059	-0.073268	0.515273	0.428592	-0.3244101	0.308094	0.256649	-0.5569333
cpunet		0.197501	0.357179		-0.013995	-0.5560234		-0.07469	0.4247139
cpu		-0.31835	-0.1595798		-0.1974525	0.8369258		-0.13081	0.7499107
luz		0.291048	0.0005574		0.1206877	-0.3029612		0.112546	0.5478098
agua		-0.07546	-0.9889679		0.2177196	0.2801287		0.299881	1.113902
calcamento					0.0578843	0.0483158		0.056441	0.0515804
mres		-0.20687	-0.2886507		0.0061268	-0.1191651		-0.04309	0.3035431
mres2		0.031615	0.0331277		0.0087001	0.0092137		0.013156	0.0121557
luzagua			0.5740318			-0.1735558			-0.1412446
luzcpu						-0.8514008			-0.8730179
aguacpu			-0.1756071			-0.1945825			-0.0085038
luzcpunet						0.6425023			-0.2751676
aguacpunet			-0.1835205			-0.0828768			-0.2250381
luznres			-0.005605			0.1444565			-0.3535664
aguanres			0.0896672			-0.0209794			0.0100216
luzeducmae			0.0247903			-0.1417206			0.3430418
agueducmae			0.0003303			-0.1080343			-0.2173954
luzeducpai			-0.6478602			-0.0240979			-0.051073
agueducpai			-0.0041314			-0.0753313			-0.1123155
redelocal			-0.290743			-23.1228			
luzlocal			-0.0761866			-23.1228			
luzrede			0.622757			0.2074821			1.053384
agualocal			0.0894055			-2.282852			
aguarede			0.0851739			0.5610045			-0.2463845



	2001																	
	sem infra			infra			infra + interações			sem infra			infra			infra + interações		
	2	4ª série EF	infra + interações	sem infra	8ª série EF	infra + interações	sem infra	3ª série EM	infra	infra + interações	sem infra	infra	infra + interações					
comp	-0.0407241	-0.04222	-0.040803	-0.03137	-0.0267533	-0.0254316	-0.01751	-0.01445	-0.0134352									
conshidra	0.016375	0.017024	0.0206386	0.006769	-0.00016	-0.0023094	0.011447	0.041511	0.0411512									
conseletr	0.0034436	-0.0116	-0.0184608	0.077904	0.0876875	0.0890982	0.032909	0.007149	0.0044756									
ilumina	-0.083893	-0.10321	-0.1023606	0.273017	0.2808111	0.2780377	0.048273	0.071219	0.061366									
local	-0.1691581	-0.19283	0.6652124	0.559479	0.7284059	-18.93142												
rede	0.8851182	0.843564	-1.650396	0.734865	0.5790112	0.1099083	0.597988	0.529923	-0.6972623									
cpunet		0.278071	0.3000475		-0.0794309	-0.419894		-0.18676	-0.4228099									
cpu		-0.17806	-0.11110334		-0.2704402	1.379579		-0.15885	0.8821683									
luz		0.400726	-1.768462		0.2997029	-0.3367437		0.156656	0.1656037									
agua		-0.05589	-0.9585632		0.489047	1.017775		0.368168	0.927254									
calcamento					0.2533514	0.2424769		0.216349	0.2039499									
nres		-0.34028	-0.5944428		0.0168195	0.0137419		-0.01236	0.0088196									
nres2		0.049944	0.0531852		0.0111737	0.0101127		0.013699	0.0134692									
luzagua			0.1836232			0.1869904			-0.4926142									
luzcpu						-1.944928			-1.222825									
aguacpu			-0.0655682			0.2856629			0.1809756									
luzcpunet						0.5956274			0.557212									
aguacpunet			-0.0201377			-0.2503654			-0.3049286									
luznres			0.0351048			0.1134733			-0.0748703									
aguanres			0.2340611			-0.1109773			0.0589678									
luzeducmae			0.1968095			-0.0318285			-0.0227643									
aguaueducmae			-0.0343177			-0.1496655			-0.1882494									
luzeducpai			-0.9293661			-0.025759			0.0367866									
aguaueducpai			-0.0798318			-0.0779201			-0.3229811									
redelocal			-0.9145759															
luzlocal			-0.0440076			-0.4073612			0.7877673									
luzrede			2.544917			0.3659247												
aguainlocal			0.1706061			20.23204												
aguarede			-0.0469539			0.1042532			0.4581795									

	2003											
	sem infra			sem infra			sem infra			sem infra		
	infra	infra + interações	4ª série EF	infra	infra + interações	8ª série EF	infra	infra + interações	3ª série EM	infra	infra + interações	
1												
comp	-0.0372585	-0.03252	-0.0286943	-0.03776	-0.0385333	-0.0349168	-0.06385	-0.06046	-0.0611069			
conshidra	-0.0304305	-0.01844	-0.0295623	-0.0069	-0.0164326	-0.0195661	0.108268	0.09996	0.0957292			
consetr	0.1997087	0.177117	0.1873378	0.022628	0.0259878	0.0278848	-0.10378	-0.10633	-0.1068132			
ilumina	0.0676042	0.05427	0.0631854	-0.0397	-0.0293827	-0.0269016	0.198249	0.234994	0.2473545			
local	-0.3253	-0.30779	-19.59662									
rede	0.829187	0.718611	1.822328	0.479051	0.4178267	-0.0434164	0.309265	0.260453	-0.6368505			
reg_metr	-0.0687088	-0.0838		-0.13454	-0.123344		-0.08349	-0.09448				
cpunet		-0.1694	-1.370621		-0.0941098	2.502492		-0.19113	-1.502723			
cpu		-0.08221	1.121768		-0.0748739	-1.238873		0.066022	0.9875214			
luz		-0.17264	-1.467235		0.0122077	-2.51946		0.863725	-0.1700007			
agua		0.001949	-0.9268672		0.465148	1.548643		-0.15128	-0.5836952			
mres		0.370697	-0.0420177		-0.0266676	-0.1453994		0.031881	-0.2055099			
mres2		-0.0226157	-0.0226157		0.0063403	0.0058091		0.002893	0.0033483			
luzagua			0.3234685			0.4195077			0.111512			
luzcpu			-1.663091			0.6547234			-1.112032			
aguacpu			0.4543671			0.5188554			0.1894237			
luzcpunet			2.012661			-1.2038			1.363232			
aguacpunet			-0.8094888			-1.400319			-0.0432716			
luzmres			0.2716242			0.1850183			0.0877742			
aguamres			0.140499			-0.0601846			0.1504228			
luzeducmae			0.0074349			0.24686			0.1349588			
agueducmae			0.059824			0.0074316			-0.0191503			
luzeducpai			0.136625			-0.0725413			-0.0802461			
agueducpai			-0.1455285			-0.1455982			-0.1387863			
redereg_metr			-0.0952461			-0.1007797			-0.246904			
redelocal			19.09763									
luzllocal			0.725088									
luzrede			-1.375696			1.123088			0.8720046			
agualocal			-0.4439656									
aguarede			0.2943755			-0.6018519			0.1709728			

	2003								
	sem infra		infra		sem infra				
	infra	infra + interações	infra	infra + interações	infra	infra + interações			
2	4ª série EF		8ª série EF		3ª série EM				
comp	-0.0072909	-0.00055	0.0127621	0.029591	0.0294153	0.0366325	-0.03864	-0.03659	-0.03686
conshidra	-0.0555593	-0.0773	-0.0903468	-0.06708	-0.0668454	-0.0837493	-0.01508	-0.03105	-0.0291601
conseletr	0.1429589	0.145944	0.1453609	-0.02115	-0.0344546	-0.0196006	-0.03198	-0.00876	-0.009193
ilumina	-0.2789737	-0.27411	-0.2233672	-0.26476	-0.284562	-0.292091	-0.06029	-0.04663	-0.0340034
local	-0.4481546	-0.4036	-0.0155322						
rede	0.9405455	0.789683	18.4092	0.537534	0.5326301	-2.372002	0.465287	0.487692	-1.543339
reg_metr	-0.3165013	-0.33548		-0.15315	-0.1407186		-0.14833	-0.13215	
cpunet	0.055594	0.055594	-11.51275		-0.0404741	0.9974084		-0.18907	-1.442643
cpu	-0.38943	-0.38943	-19.69637		0.0482511	-2.158202		0.294658	0.2355762
luz	-0.08868	-0.08868	13.83052		-0.1818589	-4.318582		0.124081	-1.202932
agua	0.209484	0.209484	0.5944249		0.8079355	1.45842		0.114675	-0.4202945
mres	0.441952	0.441952	0.1613425		-0.0240128	-0.2365261		-0.09627	-0.007752
mres2	-0.03009	-0.03009	-0.0287771		0.0072515	0.0065674		0.013358	0.0135996
luzagua			1.50763			-0.2094118			0.5494765
luzcpu			-4.995493			1.594439			0.0714937
aguacpu			24.2426			0.6271155			-0.0034341
luzcpunet			19.43494			-0.3567851			0.6763935
aguacpunet			-7.77234			-0.6753426			0.5923292
luzmres			0.2152582			0.2950456			-0.215676
aguamres			0.0545374			-0.0722117			0.1240768
luzeducmae			0.3004947			0.0496672			-0.2510119
agueducmae			-0.3340651			-0.2426715			0.1828196
luzeducpai			-0.0479401			0.084569			0.4302642
agueducpai			-0.106808			0.1861192			-0.3691968
redereg_metr			-0.2133319			-0.037128			-0.3102066
redelocal			0.9667026						
luzlocal			0.1518132						
luzrede			-17.27795			2.639982			2.032707
aguarede			-1.67097						
aguarede			-0.2373196			0.3217526			0.1750372

## *Matrícula*

Nota-se que as variáveis de infra-estrutura são significativas e aumentam a probabilidade de freqüentar escola, na maioria das regressões. Em relação ao acesso do serviço público, água e luz elétrica aumentam a probabilidade de matrícula. Para a comunicação, verificamos o efeito na mesma direção, sendo o telefone o de maior impacto, na maioria dos casos. A infra-estrutura privada gera impactos importantes. Casas maiores levam a maior freqüência escolar, principalmente devido ao maior nível de renda que esta variável está correlacionada. Outro aspecto é número de moradores. Considerando todos, nos leva a um resultado não esperado, em que quanto maior o número de moradores, em nível e marginalmente, gera uma redução na probabilidade de interesse. Restringindo a moradores com idade superior ou igual a 10 anos, o efeito se torna semelhante às análises multivariadas anteriores, em que, quanto um aumento eleva a probabilidade, mas marginalmente este acréscimo vai se reduzindo. Para a titularidade, observamos que quem tem terreno próprio, mas ainda paga, tem uma probabilidade maior de estar empregado, já quem mora de aluguel ou domicílio cedido tem menor probabilidade. Em relação aos outros fatores, quem mora na área urbana tem mais chance de estar matriculado e quem moram em aglomerados subnormais tem menor chance (para 2001 e 2004 que é estatisticamente significativo). Vale por fim ressaltar a variável renda per capita (rendapc) que é estatisticamente igual a zero, ou seja, diferenças na probabilidade de matrícula não são explicadas por diferenças no nível de renda.

	1999			2001			2003			2004		
	sem infra	infra	infra + interações	sem infra	infra	infra + interações	sem infra	Infra	infra + interações	sem infra	infra	infra + interações
local	-0.3565	-0.1642	0.0002	-0.4301	-0.1478	0.0869	-0.2929	-0.1186	0.1194	-0.4335	-0.0839	0.1871
setor	-0.0118	0.0611	16.2331	-0.2915	-0.2209	-0.5619	-0.1076	-0.0108	-0.6679	-0.1635	-0.1331	-0.2808
rendapc	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
d_luz2	-0.1169	-0.2763	-0.2763	-0.1703	-0.8262	-0.8262	-0.0816	-0.6705	-0.6705	-0.3721	-1.0801	-1.0801
d_luz3	-0.4033	-0.5001	-0.5001	-0.2969	-0.7336	-0.7336	-0.3541	-0.8088	-0.8088	-0.5677	-1.1377	-1.1377
agua	0.1114	0.2076	0.2076	0.1602	0.4386	0.4386	0.0813	0.8148	0.8148	0.2362	0.9917	0.9917
telefone	0.3867	0.7832	0.7832	0.5419	12.9517	12.9517	0.1510	0.6376	0.6376	0.4200	24.0198	24.0198
celular				-0.0926	-0.1017	-0.1017	0.0033	0.0021	0.0021	0.1410	0.1356	0.1356
cpu				0.4490	-17.0672	-17.0672	0.3261	3.7650	3.7650	0.3793	0.3312	0.3312
cpunet				0.3612	0.3247	0.3247	0.1839	0.1816	0.1816	0.1822	0.1745	0.1745
comodo	0.1155	0.1136	0.1136	0.1149	0.1139	0.1139	0.0330	0.0326	0.0326	0.0926	0.0928	0.0928
mres	-0.1582	-0.1146	-0.1146	-0.1107	0.0353	0.0353	-0.4943	-0.3432	-0.3432	-0.1163	0.0677	0.0677
mres2	-0.0012	-0.0020	-0.0020	-0.0093	-0.0127	-0.0127	0.0154	0.0124	0.0124	-0.0081	-0.0132	-0.0132
mres_10	0.4575	0.4556	0.4556	0.5147	0.5047	0.5047	1.0340	1.0271	1.0271	0.4801	0.4660	0.4660
mres2_10	-0.0309	-0.0306	-0.0306	-0.0321	-0.0307	-0.0307	-0.0657	-0.0646	-0.0646	-0.0286	-0.0266	-0.0266
proppago	-0.0359	-0.0349	-0.0349	0.0715	0.0706	0.0706	-0.0056	-0.0009	-0.0009	0.0595	0.0694	0.0694
proppagando	0.1485	0.1433	0.1433	0.1064	0.0927	0.0927	0.0015	0.0018	0.0018	0.1804	0.1826	0.1826
alugado	-0.2840	-0.2893	-0.2893	-0.2621	-0.2745	-0.2745	-0.1672	-0.1690	-0.1690	-0.2431	-0.2484	-0.2484
cedido	-0.2476	-0.2380	-0.2380	-0.1638	-0.1562	-0.1562	-0.1070	-0.0986	-0.0986	-0.2633	-0.2381	-0.2381
aguad_luz1		0.1516	0.1516		0.0795	0.0795		-0.2873	-0.2873		-0.1916	-0.1916
aguatелефone		-0.4232	-0.4232		0.1998	0.1998		-0.1164	-0.1164		0.1270	0.1270
telefoned_~1					-12.6161	-12.6161		-0.3817	-0.3817		-23.7386	-23.7386
aguanres		-0.0194	-0.0194		-0.0522	-0.0522		-0.0711	-0.0711		-0.0828	-0.0828
mresd_luz1		-0.0246	-0.0246		-0.0806	-0.0806		-0.0660	-0.0660		-0.0783	-0.0783
agualocal		-0.3075	-0.3075		-0.2140	-0.2140		-0.1993	-0.1993		-0.2667	-0.2667
locald_luz1		0.0406	0.0406		-0.1248	-0.1248		-0.1093	-0.1093		-0.1206	-0.1206
aguasetor		-0.0462	-0.0462		-0.0052	-0.0052		0.0237	0.0237		-0.1038	-0.1038
setord_luz1		-16.1340	-16.1340		0.3484	0.3484		0.6387	0.6387		0.2507	0.2507
aguacpu					0.4194	0.4194		0.0733	0.0733		0.0339	0.0339
d_luz1cpu					17.1209	17.1209		-3.5131	-3.5131			

### *Atraso*

Para o atraso escolar observamos que a maioria das variáveis de infra-estrutura são significativas. E apresentam a correlação parcial esperada de reduzir a probabilidade de acesso em favor de quem tem acesso. E as interações são em sua maioria não significativas.

	1999			2001			2003			2004		
	sem infra	infra	infra + interações	sem infra	infra	infra + interações	sem infra	Infra	infra + interações	sem infra	infra	infra + interações
rede												
local	0.1777	-0.0006	0.1600	0.5556	0.1794	0.1763	1.0464	0.9115	0.9132	0.5323	0.1867	0.1839
setor	0.1522	0.0799	0.7662	0.3590	0.0364	0.0302	0.1080	-0.0097	-0.2717	0.2926	0.0070	-0.1429
rendapc	0.0000	0.0000	0.0000	0.0000	0.1427	-0.8488	0.0119	0.0434	0.0755	0.2463	0.1935	15.0187
d_luz2		0.2724	0.0219		0.0000	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000
d_luz3		0.2469	0.0640		0.2893	0.1028		0.3129	0.4735		0.4432	0.7317
agua		-0.1536	-0.1613		0.3933	0.1587		0.2843	0.3406		0.3433	0.5856
telefone		-0.1962			-0.2768	-0.2802		-0.1537	-0.0149		-0.2265	-0.4961
celular					-0.3664	0.2736		-0.1680	-0.7720		-0.3066	10.2312
cpu					0.0966	0.1003		-0.0460	-0.0463		-0.0989	-0.0994
cpunet					-0.4209	-0.0290		-0.0530	-0.4614		-0.1940	0.0782
comodo		-0.0575	-0.0574		0.1034	0.1078		0.0056	0.0043		0.0553	0.0566
mres		0.0605	0.0747		-0.0998	-0.0998		-0.0512	-0.0511		-0.0927	-0.0929
mres2		0.0036	0.0034		0.3758	0.3567		-0.9628	-0.9254		0.3658	0.3030
mres2_10		-0.1973	-0.1986		-0.0125	-0.0110		0.0492	0.0484		-0.0100	-0.0080
proppago		0.0159	0.0160		-0.3571	-0.3500		1.7454	1.7435		-0.4124	-0.4039
proppagando		-0.0210	-0.0204		0.0235	0.0226		-0.1140	-0.1137		0.0269	0.0258
alugado		-0.0906	-0.0899		-0.1271	-0.1251		-0.1013	-0.1009		-0.0696	-0.0714
cedido		0.0691	0.0692		-0.2482	-0.2441		-0.1980	-0.1990		-0.0686	-0.0700
aguad_luz1		-0.0055	-0.0057		0.0700	0.0733		-0.0068	-0.0070		0.0593	0.0576
aguatелефone			-0.0203		-0.0013	0.0017		-0.0556	-0.0560		0.0196	0.0184
telefoned_~1			-0.0169		-0.2269	-0.2269			-0.0229			0.1106
aguamres			-0.1791		-0.0187	-0.0187			0.0999			-0.0401
mresd_luz1			-0.0010		-0.6213	-0.6213			0.5089			-10.5002
agualocal			-0.0129		0.0409	0.0409			-0.0213			0.0358
locald_luz1			0.0718		-0.0308	-0.0308			-0.0119			0.0130
aguasetor			-0.2228		-0.0619	-0.0619			-0.0115			-0.0816
setord_luz1			0.0439		0.0495	0.0495			0.2748			0.2185
aguacpu			-0.7288		0.2000	0.2000			-0.0553			0.0090
d_luz1cpu					0.8165	0.8165			0.0198			-14.8370
					-0.3976	-0.3976			0.0596			-0.2726
									0.3480			

### *Assiduidade*

Em relação à assiduidade, observamos que a maioria das variáveis de infra-estrutura são não significativas, com exceção do número de moradores, proppago (comparando pouco faltoso com o sem falta). Quando considerado a segunda regressão, comparando muito faltoso com o sem falta as variáveis se tornam significativas em maior número. Como esperado, elas apresentam o sinal econômico robusto. Por exemplo, quanto maior o número de moradores ou quem já tem domicílio próprio pago em terreno próprio, aumenta a chance de ser mais assíduo na escola em relação aos moradores com características diferentes dessas.



	1			2		
	sem infra	infra	infra + interações	sem infra	infra	infra + interações
rede	0.1929	0.1665	0.1652	0.9920	0.8816	0.8811
local	0.0387	0.0049	-0.0376	-0.1762	-0.2641	-0.2390
setor	-0.0813	-0.0909	-0.0289	0.1114	0.0921	0.8778
rendapc	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
d_luz2		0.0832	-0.0291		0.1586	-0.1365
d_luz3		0.4624	0.2927		0.3651	0.0869
agua		-0.0512	-0.0202		-0.0854	-0.1341
telefone		-0.0372	0.3368		-0.0928	-0.4963
celular		-0.0349	-0.0304		0.0819	0.0838
cpu		-0.0558	0.1070		0.0055	0.0773
cpunet		-0.0176	-0.0157		-0.1820	-0.1808
comodo		-0.0062	-0.0058		-0.0442	-0.0439
mres		0.0087	-0.0023		-0.0154	0.0040
mres2		0.0004	0.0012		0.0028	0.0030
mres_10		-0.1273	-0.1260		-0.1577	-0.1580
mres2_10		0.0067	0.0066		0.0120	0.0120
proppago		-0.1250	-0.1285		-0.0629	-0.0659
proppagando		0.0400	0.0399		0.0314	0.0300
alugado		0.0366	0.0365		0.1191	0.1181
cedido		-0.0623	-0.0682		0.0175	0.0114
aguad_luz1			-0.1938			-0.0885
aguatелефone			-0.0766			-0.0427
telefoned_~1			-0.2936			0.4497
aguanres			0.0201			0.0158
mresd_luz1			-0.0140			-0.0359
agualocal			0.1538			0.1579
locald_luz1			-0.0552			-0.1296
aguasetor			-0.0073			0.0419
setord_luz1			-0.0576			-0.8252
aguacpu			-0.1627			-0.0712

O Centro de Políticas Sociais da FGV disponibiliza site com um software amigável chamado Espelho baseado dessas regressões multivariadas, onde cada pessoa pode inserir a combinação de dados, como sexo, idade, local de moradia entre outros e em particular variáveis de infraestrutura e cenarizar por exemplo qual a probabilidade de se estar matriculado ou atrasado na escola ou a sua respectiva proficiência escolar. Posteriormente, rodamos outros modelos mais restritos que empilham todas as pesquisas e permitem que a variável ano atue apenas no coeficiente linear das regressões estes modelos apresentam tipicamente resultados mais consistentes para efeito de simulação, reduzindo as flutuações das estimativas entre anos.

Na análise empírica empreendida neste trabalho estudamos as correlações brutas e as parciais entre as variáveis de infraestrutura e as de desempenho escolar. A dificuldade central destas análises para fins de desenho de políticas é a questão da direção de causalidade entre variáveis exógenas e as endógenas. Neste sentido, uma extensão natural da análise é uma análise de diferença em diferença tomando como base a ocorrência de experiências onde algumas comunidades são beneficiadas com aumento da oferta de infraestrutura e outras não. Uma primeira candidata a experimento seria o programa Favela-Bairro no Rio onde algumas comunidades de baixa renda são beneficiados por melhoras maciças de infraestrutura privada (acesso a esgoto, luz, água e melhora da construção) e pública (pavimentação das ruas, iluminação pública, construção de praças) enquanto os críticos do programa enfatizam o fato de que outros aspectos como investimento em capital humano não foi diretamente beneficiados. Neste sentido o Favela-Bairro constituiria num experimento útil na identificação dos impactos de investimentos em infraestrutura sobre as variáveis de desempenho escolar. Esta análise seria feita através da comparação antes e depois do programa de comunidades afetadas e as não afetadas a partir dos dados do Censo 1991 e 2000. A datação do Censo e a possibilidade de abertura inframunicipal dos dados permitiriam uma análise similar as da PNAD.

#### *Análise de Diferenças em Diferenças*

*Através de “fotografias” tiradas a partir de diferentes pesquisas, poderemos captar o impacto diferenciadas de mudanças na oferta de infraestrutura, comparando pesquisas antes e depois (leia-se no caso da ECINF 1997 e 2003) de grupos beneficiados e os não beneficiados por mudanças de infraestrutura. O método estatístico que será utilizado para fazer essa avaliação será o de diferenças em diferenças. O método procura determinar o impacto de uma política exógena, utilizando dados de antes da implementação dessa política e dados depois de sua implementação.*

*Em economia, muitas pesquisas são feitas analisando os chamados experimentos naturais. Nas palavras de Wooldridge (2003), os experimentos naturais ocorrem quando algum evento exógeno, geralmente, geralmente uma mudança de política do governo, muda o ambiente no qual indivíduos, famílias, firmas ou cidades operam. Para analisarmos um experimento natural sempre temos que ter um grupo de controle que não foi afetado com a mudança e um grupo de tratamento que foi o afetado pelo evento. Ao contrário de um experimento real em que os grupos de tratamento e controle são escolhidos aleatoriamente para impedir vies nas estimativas, os grupos em um experimento natural surgem da forma em que a mudança é efetuada. Para estudarmos as diferenças entre os dois grupos precisamos de dados de antes do evento e de depois para os dois grupos. Assim nossa amostra é dividida em quatro grupos: o grupo de controle antes da mudança, o grupo de controle depois da mudança, o grupo de tratamento antes da mudança e o grupo de tratamento depois da mudança.*

*Matematicamente, podemos representar o método de diferenças em diferenças com a seguinte equação:*

$$g3 = (\bar{Y}_{2,B} - \bar{Y}_{2,A}) - (\bar{Y}_{1,B} - \bar{Y}_{1,A})$$

Onde cada  $\bar{Y}$  representa a média da variável estudada para cada ano e grupo, com o número subscrito representando o período da amostra (1, para antes da mudança e 2, para depois da mudança) e a letra representando o grupo a qual o dado pertence (A, para o grupo de controle e B, para o grupo de tratamento). E  $g_3$  será nossa estimativa a partir da diferenças em diferenças. Obtendo  $g_3$  determinamos o impacto do experimento natural sobre a variável que gostaríamos de explicar.

Representando o método através de uma regressão e criando as variáveis indicadoras (ou dummies):  $d_B$ , igual a um para os indivíduos do grupo de tratamento e zero para os indivíduos do grupo de controle; e  $d_2$ , igual a um quando os dados se referem ao segundo período, pós mudança, e zero caso os dados se referiram ao período pré-mudança; temos:

$$Y = g_0 + g_1*d_2 + g_2*d_B + g_3*d_2*d_B + \text{outros fatores}$$

Onde  $Y$  representa a variável estudada,  $g_1$  o impacto de se estar no segundo período sobre a variável estudada,  $g_2$  o impacto de se estar no grupo de tratamento sobre a variável estudada, e  $g_3$  o impacto pós evento do grupo de controle sobre a variável estudada (que é justamente o que se quer descobrir). Assim,  $g_0$  capta justamente o valor esperado da variável estudada quando se analisa o grupo de controle antes da mudança, o que nos dá, basicamente, o parâmetro de comparação.

Porém é preciso controlar por outros fatores relevantes na regressão, o que no jargão econométrico quer dizer que, antes de alegarmos que  $g_3$  nos dará o impacto da política exógena, temos que descobrir e isolar o efeito de todas as outras variáveis que podem estar causando mudanças na variável estudada. Isso é feito inserindo as variáveis de controle relevantes na regressão, como foi mostrado na segunda equação, assim evita-se que efeitos de outras variáveis produzam viés na nossa estimação e determinamos o efeito puro do experimento natural sobre a variável que gostaríamos de explicar.

## **Conclusões**

Este relatório avalia as correlações entre acesso à infra-estrutura e indicadores de desempenho escolar tais como matrícula, frequência, repetência e proficiência escolar. As variáveis de infra-estrutura utilizadas são acesso a serviços públicos (abastecimento de água, esgotamento sanitário e iluminação elétrica e comunicações.) e infra-estrutura privada (número de pessoas no domicílio, direito de propriedade). A pesquisa consiste na geração, descrição e análise de um conjunto de padrões de correlações brutas e parciais a partir de duas bases de microdados. Utilizamos pesquisa domiciliar a PNAD para uma avaliação das correlações entre as condições de acesso à infra-estrutura doméstica e algumas variáveis sociais. Os efeitos sociais da provisão de infra-estrutura em casa e aqui também nas escolas serão captados através de informações do Ministério da Educação por meio de avaliação de desempenho do SAEB.

## **Resultados da Análise Bivariada**

### **Proficiência e Reprovação - SAEB**

Apresentamos a relação das variáveis de infra-estrutura tomadas isoladamente na proficiência - exame de Matemática - e reprovação. Vale destacar que, na análise da variável reprovação (referente à questão “Você já foi reprovado?”), esta apresenta os seguintes valores (nenhuma reprovação, uma reprovação e duas ou mais reprovações). Segue a análise por grupos de variáveis de infra-estrutura: comunicações (acesso à internet, computador etc), serviços públicos (acesso à luz, água etc).

## **Comunicações**

As conseqüências do acesso a computadores em casa são relativamente desconhecidas, e particularmente, a literatura dos impactos educacionais de computadores em casa é escassa. A importância dessa análise reside na evidência encontrada na literatura que seu uso em casa pode compensar a falta da tecnologia da informação na escola. Assim observa-se uma proficiência maior para os alunos que tem acesso à Internet em casa, via computador, para todos os anos, tendo um diferencial maior em relação aos que não tem, no ano mais recente. Discriminado por série este impacto nota-se um diferencial crescente em relação a séries mais altas. Ou seja, a diferença da nota entre quem tem acesso e não tem é maior para quem está na 3ª série do Ensino Médio, do quem está na 8ª série do Ensino Fundamental, que é maior em relação a quem está na 4ª série do Ensino Fundamental. Mesmo considerando os que tem computador em casa, sem discriminar se tem acesso à internet ou não, o desempenho foi melhor para aqueles que possuem.

Observa-se que quanto maior o número de computadores disponíveis para os alunos melhor o desempenho dos mesmos na escola. Uma estatística exclusiva para o ano de 2001 que inferimos é se o aluno usa o computador para fazer a lição de casa. O que notou-se é que não há uma relação muito direta entre o uso do computador para fazer a lição e a proficiência dos alunos. Apenas no caso do aluno nunca utilizar, este têm uma nota, em média, menor em relação aos que usam, principalmente para a 8ª série e a 3ª série.

Corroborando o resultado apresentado para a proficiência, observa-se uma maior repetência para quem não tem computador com acesso à internet. Em todas séries, observa-se o mesmo resultado. Bem como para análise na escola, o maior acesso a computadores reduz a média de repetência entre os alunos, principalmente em 2003. Em relação ao uso de computadores na realização de tarefas escolares, evidenciou-se que o uso do computador para fazer tarefas escolares pode reduzir a reprovação e até melhorar o desempenho do aluno, mas desde que o mesmo não se utilize deste artifício sempre.

### **Serviços Públicos**

O acesso à infra-estrutura não tem sido objeto de estudo explícito na literatura. Apenas tem servido como controle para se avaliar o efeito de algum fator em alguma variável educacional (performance, matrícula etc). Então torna-se imprescindível focar a análise em aspectos de acesso à infra-estrutura e o desempenho escolar. A questão que surge (e será dirimida na análise multivariada) é o quanto que o acesso à luz, água etc e melhora a proficiência escolar, tendo-se controlado para outros fatores (como educação dos pais, unidade federativa, número de residentes na mesma casa etc). Observa-se que o uso da eletricidade e da água tem um correlação positiva com a proficiência e há um peso mais negativo sobre quem não tem acesso à eletricidade do que à água. Analisamos também algumas variáveis de infra-estrutura ligadas à escola. Observamos que a conservação das instalações hidráulicas e elétricas tem uma correlação positiva com o desempenho dos alunos. Outro aspecto importante é a iluminação no local de estudo, mostrando um melhor desempenho educacional nas salas iluminadas.

O acesso à infra-estrutura também reduz o índice de reprovação, mas ao contrário do desempenho escolar, a eletricidade não gera um impacto mais negativo do que o acesso à água, em 2001 e 2003. Apesar da falta de robustez em relação à infra-estrutura escolar, geralmente colégios com instalações mais adequadas apresentam menor índice de repetência.

### **Outros fatores**

Um dos elementos candidatos a explicar este diferencial de desempenho seria a diferença de acesso a infraestrutura entre áreas rural e urbana. Observa-se uma pequena evolução de 2001 para 2003 nas notas dos dois tipos de escola, sendo 1.83% da rural e 1.2% da urbana. O índice de reprovação para escolas urbanas é bem menor do que as rurais, principalmente para as rurais.

Outra deficiência que se apresenta no país que citamos brevemente é a da escola pública em relação à particular, como é notado nos valores de proficiência observado abaixo. As duas redes de ensino evoluíram pouco de 2001 para 2003 (em torno de 0.8%). O índice de reprovação para todas as séries é bem maior nas escolas públicas do que privadas, apesar de que, de 2001 para 2003 as reprovações caíram bastante na rede pública de ensino.

### **Matrícula, Atraso e Assiduidade - PNAD**

A análise bivariada estende-se agora à matrícula, atraso e assiduidade. As variáveis de infra-estrutura encontradas na PNAD diferem um pouco das encontradas no SAEB, vistas na análise precedente. Aqui, restringimos a população de interesse à faixa etária de 5 à 17 anos que estão na “idade escolar ativa”.

### **Serviços Públicos**

Em relação aos serviços públicos, notamos que o acesso à água gera maiores índices de matrícula. E a luz recebida na forma elétrica também incentiva à frequência escolar. Além disso, no contexto dinâmico, a média de matrícula escolar tem crescido entre aqueles que tem acesso à água e luz elétrica, ao longo de todos os anos. Para a variável atraso, temos a mesma dinâmica que a apresentada para a matrícula, em favorecimento a quem tem acesso à água e luz elétrica. E tem também decaído ao longo dos anos. Relembrando que para assiduidade, observamos um resultado abaixo contraditório àquelas discutidos mais acima, no sentido de que quem tem acesso a estes serviços públicos a frequência escolar é menor..

### **Comunicações**

Continuando a análise das conseqüências do uso do computador na educação, verificamos, a priori, que o fato de possuir este bem em casa gera maiores incentivos à matrícula escolar, mas que a diferença em relação a quem não tem, vem decrescendo ao longo dos anos, mas permanece significativa. Em relação ao uso da internet, observamos o mesmo comportamento e trajetória. Para a variável atraso, observamos que a média de atraso é maior entre aqueles que não tem computador em casa e nem acesso à internet. O mesmo comportamento observamos para aqueles que não tem telefone fixo e/ou celular. Em relação à assiduidade, observamos que quem tem computador e internet, bem como telefonia fixa são mais assíduos. Apenas quem tem celular parece faltar mais à escola.

### **Infra-estrutura privada e outros fatores**

Em relação, primeiramente, ao número de cômodos nota-se um relação crescente com matrícula, assiduidade e atraso escolar. A variável “local” é específica para o local onde os indivíduos moram. E corrobora-se os esperado, o qual, pessoas que moram em centros urbanos tem um maior índice de matrícula, de assiduidade e menores índices de atraso escolar em relação aos que moram na zona rural.

### **Resultados da Análise Multivariada**

O desempenho dos alunos pode ser afetado pelo acesso a serviços públicos como luz, água etc. Melhor infra-estrutura, tanto na casa como na escola, deve melhorar a produtividade dos estudantes, reduzindo portanto, a repetência. Mas a questão que se levanta aqui é que outros fatores podem também afetar estas variáveis de interesse. Assim, os controles assumidos ajudam a isolar de forma mais precisa o efeito da infra-estrutura no desempenho escolar. Assim, por exemplo, a educação dos pais pode afetar positivamente a produtividade de seus filhos, mesmo que haja certa deficiência em infra-estrutura na escola. Além disso, o número de moradores na casa do aluno pode também influenciar. Temos observado que há geralmente um número ótimo de moradores que ajudam a melhorar a performance escolar do estudante. Além disso, controlamos também por sexo, cor e Unidade Federativa (UF), visto que as características escolares podem divergir entre esses grupos.

### ***Proficiência***

As regressões para a equação de proficiência foram estimadas por série. O modelo estimado foi um modelo linear nas co-variáveis. Estimamos a regressão por mínimos quadrados ordinários, mas consideramos para a hipótese de heterocedasticidade dos erros. Assim, estimamos a matriz de variância através do estimador de White (1980) produzindo erros padrões robustos a heterocedasticidade.

A partir das regressões por mínimos quadrados ordinários de variáveis contínuas, observa-se que a falta de eletricidade tem um efeito mais perverso que a água, na grande maioria dos casos. O impacto da infra-estrutura escolar (conservação das instalações hidráulicas e elétricas e iluminação na sala de aula) também afetam positivamente a proficiência escolar. O acesso a computadores pode desenvolver no aluno uma melhor “aptidão escolar”, gerando bons resultados na escola. O acesso à internet em casa não apresenta uma direção clara, mas parece que em alguns casos pode gerar uma queda da produtividade do aluno. Quando comparamos o impacto de se ter computador em casa com o impacto de se ter computador na escola disponível para o aluno, o primeiro fator tem peso, em termos de magnitude, bem maior que o segundo. Isso provavelmente é causado pelo melhor ambiente encontrado em casa frente à escola, e pelo fato do aluno ter uma menor competição pelo uso em casa em relação à escola. Logo, em termos gerais, a infra-estrutura pode impactar positivamente na proficiência escolar, podendo até chegar a um aumento em torno de 12 pontos na nota (no caso de se ter computador em casa, em relação a quem não tem, *ceteris paribus*). Em relação à infra-estrutura privada, observamos o comportamento esperado: um efeito positivo em nível, e efeito negativo no termo quadrático.

### ***Reprovação***

A título de exemplificação, interpretando a variável nres no ano de 1999, temos que, um morador a mais no domicílio aumenta o log da razão por 0.141289 (comparando reprovado uma vez com nenhuma vez), ou seja, reduz-se a probabilidade de não ser reprovado e/ou aumenta-se a probabilidade de ser reprovado uma vez. Das variáveis de infra-estrutura, observamos que a maioria é estatisticamente não significativa. Somente cpu, cpunet, e água, nres e nres2 em alguns casos são significativas. Destas somente o acesso a computador e internet apresentam o sinal negativo e de acordo com o esperado (com exceção de 1999), ou seja, quem tem computador e internet em casa, tem uma maior chance

de não ser reprovado e uma menor chance de ser reprovado. As interações na grande maioria não são estatisticamente válidas e sua inclusão reduz o efeito das outras e/ou tornando-as também não significativas, principalmente na 8ª série e na 3ª série.

### *Matrícula*

Nota-se que as variáveis de infra-estrutura são significativas e aumentam a probabilidade de freqüentar escola, na maioria das regressões. Em relação ao acesso do serviço público, água e luz elétrica aumentam a probabilidade de matrícula. Para a comunicação, verificamos o efeito na mesma direção, sendo o telefone o de maior impacto, na maioria dos casos. A infra-estrutura privada gera impactos importantes. Casas maiores levam a maior freqüência escolar, principalmente devido ao maior nível de renda que esta variável está correlacionada. Outro aspecto é número de moradores. Considerando todos, nos leva a um resultado não esperado, em que quanto maior o número de moradores, em nível e marginalmente, gera uma redução na probabilidade de interesse. Restringindo a moradores com idade superior ou igual a 10 anos, o efeito se torna semelhante às análises multivariadas anteriores, em que, quanto um aumento eleva a probabilidade, mas marginalmente este acréscimo vai se reduzindo. Para a titularidade, observamos que quem tem terreno próprio, mas ainda paga, tem uma probabilidade maior de estar empregado, já quem mora de aluguel ou domicílio cedido tem menor probabilidade. Em relação aos outros fatores, quem mora na área urbana tem mais chance de estar matriculado e quem moram em aglomerados subnormais tem menor chance (para 2001 e 2004 que é estatisticamente significativo). Vale por fim ressaltar a variável renda per capita (rendapc) que é estatisticamente igual a zero, ou seja, diferenças na probabilidade de matrícula não são explicadas por diferenças no nível de renda.

### *Atraso*

Para o atraso escolar observamos que a maioria das variáveis de infra-estrutura são significativas. E apresentam a correlação parcial esperada de reduzir a probabilidade de acesso em favor de quem tem acesso. E as interações são em sua maioria não significativas.

### *Assiduidade*

Em relação à assiduidade, observamos que a maioria das variáveis de infra-estrutura são não significativas, com exceção do número de moradores, proppago (comparando pouco faltoso com o sem falta). Quando considerado a segunda regressão, comparando muito faltoso com o sem falta as variáveis se tornam significativas em maior número. Como esperado, elas apresentam o sinal econômico robusto. Por exemplo, quanto maior o número de moradores ou quem já tem domicílio próprio pago em terreno próprio, aumenta a chance de ser mais assíduo na escola em relação aos moradores com características diferentes dessas.

## **Simulador e extensões**

O Centro de Políticas Sociais da FGV disponibiliza site com um software amigável chamado Espelho baseado dessas regressões multivariadas, onde cada pessoa pode inserir a combinação de dados, como sexo, idade, local de moradia entre outros e em particular variáveis de infraestrutura e cenarizar por exemplo qual a probabilidade de se estar matriculado ou atrasado na escola ou a sua respectiva proficiência escolar. Posteriormente, rodamos outros modelos mais restritos que empilham todas as pesquisas e permitem que a variável ano atue apenas no coeficiente linear das regressões estes modelos apresentam

tipicamente resultados mais consistentes para efeito de simulação, reduzindo as flutuações das estimativas entre anos.

Na análise empírica empreendida neste trabalho estudamos as correlações brutas e as parciais entre as variáveis de infraestrutura e as de desempenho escolar. A dificuldade central destas análises para fins de desenho de políticas é a determinação da direção de causalidade entre variáveis exógenas e as endógenas. Neste sentido, uma extensão natural é a análise de diferença em diferença tomando como base a ocorrência de experiências onde algumas comunidades são beneficiadas com aumento da oferta de infraestrutura e outras não. Uma primeira candidata a experimento seria o programa Favela-Bairro no Rio onde algumas comunidades de baixa renda foram beneficiadas por melhoras maciças localizadas de infra-estrutura privada e pública (acesso a esgoto, luz, água e melhora da construção, pavimentação das ruas, iluminação pública e construção de praças) enquanto os críticos do programa enfatizam o fato de que outros aspectos como investimento em capital humano não foi diretamente beneficiados. Neste sentido o Favela-Bairro constituiria num experimento útil na identificação dos impactos de investimentos em infraestrutura sobre as variáveis de desempenho escolar. Esta análise seria feita através da comparação antes e depois do programa de comunidades afetadas e as não afetadas a partir dos dados do Censo 1991 e 2000. A datação do Censo e a possibilidade de abertura inframunicipal dos dados permitiriam uma análise similar as da PNAD, mas com a possibilidade de identificação da relação de causalidade entre a provisão de infra-estrutura e a obtenção de avanços nos indicadores educacionais.

### **Referências Bibliográficas**

- Agresti, Alan An Introduction to Categorical Data Analysis, New York :Wiley Series in Probability and Statistics, 1996.
- Fairlie, Robert W. The Effects of Home Computers on School Enrollment 2003
- Heckman, J. The Lessons from The Technology of Skill Formation, Working Paper 11142, Fev 2005
- Selwyn, Neil The effect of using a home computer on students' educational use of IT Computers & Education 31 (1998) 211-227
- White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica* 48, 817-838.
- Wooldridge, Jeffrey M. Introductory econometrics: a modern approach. Cincinnati: South-Western College Publishing, 2003



## Apêndice

2001

-> serie = 4

Você tem em sua casa acesso à internet?	média(profic)	N(profic)
Sim	185.7531	621,621
Não	177.9812	2734229

-> serie = 8

Você tem em sua casa acesso à internet?	média(profic)	N(profic)
Sim	274.4204	519,769
Não	237.9273	2403585

-> serie = 11

Você tem em sua casa acesso à internet?	média(profic)	N(profic)
Sim	311.3551	466,185
Não	267.378	1554201

2003

-> serie = 4

na sua casa tem computador com internet?	média(profic)	N(profic)
Sim	208.6114	471,070
Não	173.2101	3179690
não sei	155.3806	71,494.60

-> serie = 8

na sua casa tem computador com internet?	média(profic)	N(profic)
Sim	284.2535	567,048
Não	236.7239	2591274
não sei	222.2054	22,599.40

-> serie = 11

na sua casa tem computador com internet?	média(profic)	N(profic)
sim	318.9454	502,285
não	266.4105	1596230
não sei	250.1846	7,884.80

2001

-> serie = 4

onde você mora existe eletricidade?	média(profic)	N(profic)
sim	178.7208	3430001
nao	151.8829	56,097.90

-> serie = 8

onde você mora existe eletricidade?	média(profic)	N(profic)
sim	244.5079	2919663
nao	206.4126	46,910.60

-> serie = 11

onde você mora existe eletricidade?	média(profic)	N(profic)
sim	277.2634	2034147
nao	243.8768	13,949.90

-> serie = 4

onde você mora chega água pela torneira?	média(profic)	N(profic)
sim	180.1282	2963006
nao	164.6399	527,508

-> serie = 8

onde você mora chega água pela torneira?	média(profic)	N(profic)
sim	245.321	2773006
nao	221.489	192,582

-> serie = 11

onde você mora chega água pela torneira?	média(profic)	N(profic)
sim	277.9372	1961199
nao	255.753	88,567.80

2003

-> serie = 4

onde você mora existe eletricidade?	média(profic)	N(profic)
sim	179.0905	3536947
nao	146.2375	209,499

-> serie = 8

onde você mora existe eletricidade?	média(profic)	N(profic)
sim	245.6929	3119823
nao	213.5647	62,049.90

-> serie = 11

onde você mora existe eletricidade?	média(profic)	N(profic)
sim	279.2073	2084068
nao	248.5412	24,049.80

-> serie = 4

onde você mora chega água pela torneira?	média(profic)	N(profic)
sim	179.8557	3428302
nao	149.1777	320,539

-> serie = 8

onde você mora chega água pela torneira?	média(profic)	N(profic)
--	---------------	-----------

sim	247.0163	2981482
nao	218.4007	191,287

-> serie = 11

onde você mora chega água pela torneira?	média(profic)	N(profic)
sim	280.0258	2017144
nao	253.2893	89,681.30