

# **TEXTO PARA DISCUSSÃO Nº 1298**

## **EDUCAÇÃO: UM ESCUDO CONTRA O HOMICÍDIO?**

**Sergei Suarez Dillon Soares**

Brasília, agosto de 2007



# **TEXTO PARA DISCUSSÃO Nº 1298**

## **EDUCAÇÃO: UM ESCUDO CONTRA O HOMICÍDIO?\***

**Sergei Suarez Dillon Soares\*\***

Brasília, agosto de 2007

---

\* Agradeço a Gláucio Ary Dillon Soares pelo incentivo a escrever este texto, a Helder Ferreira e Herton Araújo pela ajuda no uso dos dados do Sistema de Informações sobre Mortalidade, e a Rafael Ribas por sugestões úteis no que concerne à imputação de dados omitidos. Agradeço também a Helder Ferreira, Natália Fontoura e Daniel Cerqueira por comentários que muito enriqueceram o texto. Todos os erros remanescentes e interpretações são de responsabilidade exclusiva do autor. Toda a programação usada neste texto está disponível e pode ser fornecida por e-mail (sergei.soares@ipea.gov.br).

\*\* Técnico de Planejamento e Pesquisa da Diretoria de Estudos Sociais do Ipea.

## **Governo Federal**

### **Secretaria de Planejamento de Longo Prazo da Presidência da República**

**Ministro** – Roberto Mangabeira Unger



Fundação pública vinculada à Secretaria de Planejamento de Longo Prazo da Presidência da República, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiro – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

#### **Presidente**

Marcio Pochmann

#### **Diretora de Estudos Sociais**

Anna Maria T. Medeiros Peliano

#### **Diretora de Administração e Finanças**

Cinara Maria Fonseca de Lima

#### **Diretor de Estudos Setoriais**

João Alberto De Negri

#### **Diretor de Estudos Regionais e Urbanos**

José Aroudo Mota (substituto)

#### **Diretor de Estudos Macroeconômicos**

Paulo Mansur Levy

#### **Diretor de Cooperação e Desenvolvimento**

Renato Lóes Moreira (substituto)

#### **Chefe de Gabinete**

Persio Marco Antonio Davison

#### **Assessor-Chefe de Comunicação**

Murilo Lôbo

URL: <http://www.ipea.gov.br>

Ouvidoria: <http://www.ipea.gov.br/ouvidoria>

ISSN 1415-4765

JEL K42

## **TEXTO PARA DISCUSSÃO**

Publicação cujo objetivo é divulgar resultados de estudos direta ou indiretamente desenvolvidos pelo Ipea, os quais, por sua relevância, levam informações para profissionais especializados e estabelecem um espaço para sugestões.

As opiniões emitidas nesta publicação são de exclusiva e de inteira responsabilidade do(s) autor(es), não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou da Secretaria de Planejamento de Longo Prazo da Presidência da República.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

# SUMÁRIO

SINOPSE

ABSTRACT

1 INTRODUÇÃO	7
2 REVISÃO BIBLIOGRÁFICA – PERPETRADORES NOS PAÍSES ANGLÓFONOS, VÍTIMAS NO BRASIL	8
3 DADOS E METODOLOGIA	13
4 RESULTADOS	19
5 DISCUSSÃO E CONCLUSÃO	26
REFERÊNCIAS	30
ANEXOS	32



## **SINOPSE**

O objetivo deste texto é fazer uma análise preliminar das relações entre nível de instrução formal e probabilidade de ser vítima de homicídio. Para tanto, são utilizados dados do Sistema de Informações sobre Mortalidade do Ministério da Saúde (SIM/MS) de 1999 a 2004, do Censo Demográfico de 2000, e das Pesquisas Nacionais por Amostra de Domicílios do Instituto Brasileiro de Geografia e Estatística (Pnads/IBGE) de 1999 a 2004 – exceto 2000, quando a Pnad não foi a campo.

O texto consiste tanto de uma análise exploratória que compara as taxas de mortalidade por homicídio por idade para diferentes níveis de instrução formal, quanto de uma análise de regressão para achar coeficientes de correlação parciais. As regressões são estimadas com a utilização de dados agrupados e pareados por células definidas por sexo, idade, região de residência, cor e escolaridade. Estimou-se um modelo linear de probabilidade de uma regressão logística.

O texto reforça três resultados já conhecidos. O primeiro é que mulheres sofrem de homicídio a uma taxa quase dez vezes inferior à dos homens. O segundo, que negros têm maior probabilidade de morrer vítimas de homicídio que brancos. O terceiro é que jovens entre 16 e 36 anos perfazem o grande grupo de risco para a morte por homicídio.

O principal resultado inovador é que a escolaridade reduz significativamente o risco de morte por homicídio, embora sua magnitude mais exata dependa do modelo estimado.

As correlações encontradas, no entanto, não podem ser usadas para inferir causalidade sem a devida cautela. Há certamente variáveis omitidas, exercendo influência tanto sobre a morte por homicídio como sobre o nível de instrução formal, que podem levar a uma correlação espúria entre os dois. Uma análise que se utilize de variáveis instrumentais para alcançar tal objetivo será objeto de um trabalho futuro. No entanto, a magnitude do efeito é tão forte, que mesmo que apenas uma fração seja de fato causal, ainda assim a educação representa uma das melhores políticas públicas para a redução da violência letal.

## **ABSTRACT**

The objective of this text is to make a preliminary analysis of the relations between educational level and victimization by homicide. To this end, I use the following data sources: the Integrated Mortality System (SIM) between 1999 and 2004; the Demographic Census of 2000 and the National Household Surveys between 1999 and 2004 (except for 2000, year in which there was no household survey).

The text consists of both an exploratory analysis comparing homicide rates per 100,000 inhabitants according to age, educational level and sex and a regression analysis to find partial correlation coefficients. The regressions are estimated using cells defined by sex, age, region of residence, skin color and schooling level. The estimation methods used were the linear probability model and logistic regression.

My results reinforce three already well-known results. The first is that women suffer homicide rates that are roughly one-tenth that of men. The second is that

negroes suffer much higher homicide rates than whites. Finally the most important homicide risk group are youths between 16 and 36.

The most important new result is the importance of educational in preventing homicide. Although the exact magnitude depends upon the model being estimated, in all of them schooling is significantly and negatively related to death by homicide.

However, caution must be exercised in interpreting the partial correlation coefficients found cannot be interpreted as causal. There are no doubt omitted variables that have a causal relation both with educational level and homicide risk and this may lead to spurious correlation between the two. An analysis using instrumental variables to separate causal and endogenous magnitudes will be the subject of an upcoming study. However, the magnitude of the effect is so strong that even if a fraction is causally due to education, schooling still is one of the most important public policies for reducing homicide.



# 1 INTRODUÇÃO

O crime violento no Brasil atinge níveis assustadores e mobiliza o debate político e a mídia. Segundo Ferreira e Araújo (2006), há 26,4 mortes por homicídio por cem mil pessoas no Brasil (2000), contra: 5,4 no Chile (2001); 5,8 na Costa Rica (2002); 10,1 no México (2001) – país considerado violento devido ao tráfico de drogas; 1,8 em Portugal (2002); e apenas um por cem mil na Espanha (2001).<sup>1</sup> Note-se que são todos países de língua espanhola ou portuguesa e que compartilham nossa cultura ibérica. Tão preocupante quanto o nível elevado das taxas é o fato de estas estarem aumentando quase que ininterruptamente há duas décadas. Em 1993, quando o Brasil já estava assustado com o aumento da violência, a taxa de homicídios era de 20,3 indivíduos por cem mil habitantes. Não há dúvida que trata-se de um problema grave para as políticas públicas no país.

Dada a importância do crime violento no Brasil, é curioso que exista relativamente pouco trabalho quantitativo por parte da comunidade acadêmica sobre o tema. Em parte, isto se deve aos sistemas de informação extremamente débeis e à pouca confiabilidade dos dados sobre criminalidade no país. Há pouca informação sistemática quando comparada a outras áreas do conhecimento, como, por exemplo, o funcionamento do mercado de trabalho, a macroeconomia ou o desempenho educacional das nossas crianças. Talvez o único crime violento sobre o qual existam informações confiáveis é o homicídio, uma vez que este aparece em uma base de dados de outra área: o Sistema de Informações sobre Mortalidade do Ministério da Saúde (SIM/MS).

Um argumento sustentado por muitos é que o crime seria uma relação social fortemente influenciada por fatores fora do âmbito do aparelho repressor do Estado.<sup>2</sup> Segundo esta linha de argumentação, a pobreza, a desigualdade, o desemprego ou a escolaridade exerceriam impactos fortes sobre as taxa de crime violento. Vários destes fatores já foram objeto de análise quantitativa na literatura, parcialmente resenhada adiante. No entanto, há muito pouco trabalho realizado sobre a relação entre educação e crime no nosso país, foco principal deste trabalho. Dado o problema estatístico explicitado, seguiremos aqui o caminho trilhado pela quase totalidade da literatura disponível, e a análise será centrada no crime melhor documentado: o homicídio.

Este texto está dividido em quatro partes, além desta introdução. Na seção que segue será feito um rápido resumo de parte da literatura sobre a relação entre escolaridade e crime, com ênfase sobre escolaridade e homicídio. A seção 3 trata das fontes de dados e da metodologia a ser usada no texto, e pode ser abstraída pelos leitores sem inclinação matemática. A seção 4 contém os resultados *per se* e, na última seção, discute-se o que este autor contempla enquanto principais conclusões.

---

1. Os resultados de Ferreira e Araújo (2006), *grasso modo*, coincidem com os de Waiselfisz (2004), talvez porque a fonte de ambos seja a mesma: a Organização Mundial da Saúde (OMS).

2. Esta posição não implica necessariamente que o crime não sofra influência do aparelho repressor do Estado, mas apenas que outras variáveis também o influenciam fortemente.

## 2 REVISÃO BIBLIOGRÁFICA – PERPETRADORES NOS PAÍSES ANGLÓFONOS, VÍTIMAS NO BRASIL

Apesar de haver relativamente pouco trabalho quantitativo relacionando nível de instrução e morte por homicídio no Brasil, algo há. Fora do Brasil, tem-se uma literatura maior, embora esteja centrada na propensão a cometer um crime (inclusive homicídio), e não a morrer vítima de homicídio. Acredito que isto seja mais uma vez decorrência da disponibilidade de dados. Há, nos Estados Unidos (EUA) e na Inglaterra, tanto registros administrativos muito melhores que no Brasil,<sup>3</sup> como também pesquisas em painel que seguem jovens ao longo de vários anos. Incluem-se nesta revisão da literatura estes trabalhos que, a rigor, respondem uma pergunta que não está sendo colocada aqui devido a uma certa identidade, embora longe de perfeita, no perfil estatístico de vítimas e perpetradores. Contudo, esta mesma identidade será evocada mais adiante, na discussão dos resultados.

De fato, embora não haja unanimidade na literatura estadunidense e inglesa a respeito da intensidade em si, todos os trabalhos outorgam à escolaridade alguma importância como fator redutor da criminalidade.

Johnson, Kantor e Fishback (2007) estudam as taxas de crime de várias cidades nos EUA durante a grande depressão, e chegam à conclusão de que um aumento na taxa de alfabetização da cidade de 1% reduz a taxa de crime contra propriedade em 0,6%. Os resultados para crime contra a pessoa não são significativos, mas é importante lembrar que os dados usados por eles são agregados por cidade, e, portanto, sofrem viés de agregação. A renda não é incluída como variável explicativa, mas a taxa de emprego da cidade é incluída como indicador de atividade econômica.

Witte e Tauchen (1994), fazendo uso de dados individuais de painel, encontram que a frequência à escola reduz a probabilidade de se cometer um crime em aproximadamente 10%, o que é um resultado muito forte, embora potencialmente endógeno. Os autores não utilizam controles de renda domiciliar *per capita*; em vez disso, utilizam-se de dados de emprego e categoria ocupacional.

Mocan e Rees (1999), trabalhando com a pesquisa em painel *National Longitudinal Study of Adolescent Health*, encontram um efeito redutor da escolaridade dos pais sobre a propensão ao crime, apesar de este se mostrar maior para jovens do sexo feminino do que para jovens do sexo masculino. Os resultados encontrados chegam a uma redução de oito pontos percentuais para alguns tipos de crime. Porém, dado que esses autores não contam com informações sobre a renda da família, é possível que parte do impacto da escolaridade seja, na verdade, devido à renda.

Imai, Katayama, e Krishna (2006) dividem os indivíduos entre aqueles que seriam propensos ao crime e os que não o seriam, e encontram um impacto redutor do crime tanto do nível escolar do criminoso potencial como de seus pais. Estes autores não usam renda como controle, embora trabalhem com uma variável que indica se a família do indivíduo se encontra ou não na pobreza. Curiosamente, o impacto é maior para os indivíduos propensos ao crime.

---

3. Os registros administrativos sobre crime são muito melhores que no Brasil, mas isso não significa dizer que sejam perfeitos. Tanto nos EUA como na Inglaterra há subdeclaração de crimes.

Grogger (1998), em um dos estudos mais cuidadosos entre os conhecidos pelo autor deste texto, estima a probabilidade de se cometer um crime como função da renda de um indivíduo. O trabalho estima a propensão ao crime em dois estágios, sendo o primeiro uma estimação independente da renda para controlar uma possível causalidade reversa. O autor também inclui nível de instrução como variável de controle, e encontra uma redução na probabilidade de se cometer um crime se o indivíduo terminou o segundo grau. No entanto, como o efeito não é significativo, Grogger argumenta que o maior impacto da educação se dá mediante seu papel de determinante da renda, e não diretamente como redutor da propensão a cometer crimes. É importante frisar que Grogger faz um controle muito metuculoso da renda.

Os estudos aqui mencionados, embora usem a escolaridade como variável explicativa, não a têm como foco central. Em um relatório especial do Departamento de Justiça dos EUA, Harlow (2003) encontra que a população carcerária tem níveis educacionais muito inferiores aos da população em geral. Os registros administrativos mostram que 41% da população que está cumprindo pena não têm diploma secundário, contra apenas 18% da população total. No outro extremo, 48% da população geral têm diploma de nível superior, contra menos que 13% da população presa. É claro que o fluxo criminal imperfeito e a ação do tempo fazem com que a população carcerária não seja uma amostra representativa da população que cometeu crimes.

Por fim, Lochner e Moretti (2001) usam registros administrativos de prisões feitas pela polícia e de indivíduos que estão cumprindo pena em penitenciárias, além de dados do painel *National Longitudinal Survey of Youth*,<sup>4</sup> para investigar o papel da escolaridade enquanto inibidora do crime. É o estudo mais cuidadoso que relaciona esses dois fatores, e seus resultados são fortes. Segundo registros administrativos penitenciários, a probabilidade de estar cumprindo pena prisional cai de 1,6% para indivíduos sem um diploma de segundo grau para 0,4% para indivíduos com diploma superior. É claro que esta diferença pode sofrer influência da capacidade que os infratores mais educados teriam de se manter fora da prisão, mesmo após cometerem um crime violento, mas a diferença é considerável. No entanto, segundo o *National Longitudinal Survey of Youth*, no qual não há este tipo de viés de seleção, não terminar o segundo grau *dobra* a probabilidade de um indivíduo ser preso<sup>5</sup> por um crime violento. Dado que os resultados podem ser endógenos – ou seja, há um terceiro fator omitido que leva tanto à baixa instrução quanto à alta criminalidade –, os autores usam mudanças nas leis de educação compulsória como instrumento, e chegam aos mesmos resultados. É particularmente relevante que quando os resultados são desagregados por tipo de crime, homicídio e lesão corporal são justamente os mais sensíveis a aumentos na escolaridade. Ademais, os resultados são tão maiores quanto mais propenso ao crime for o indivíduo – por exemplo, o impacto sobre negros é quatro vezes superior ao impacto sobre brancos. Nas palavras dos autores: “(...) o término da educação secundária reduz a atividade criminal de modo significativo.

---

4. O nome *National Longitudinal Survey of Youth* (NLSY) se refere a duas pesquisas feitas pelo Bureau of Labor Statistics, que seguem um grupo de jovens por vários anos. O primeiro grupo começou em 1979, com 12.686 jovens com idade entre 14 e 22 anos, e os seguiu até 1994. O segundo começou em 1997, com 9.000 jovens entre 12 e 16 anos, e os segue anualmente até hoje.

5. *Preso* neste texto é a tradução do inglês de *arrest*, que significa *preso pela polícia*, e não *indiciado* ou *condenado*.

Este achado é robusto e se sustenta mesmo com diferentes estratégias de identificação e diferentes definições do que é o crime.”<sup>6</sup>

No Brasil, há uma extensa literatura e tradição de pesquisa usando como fonte o Sistema de Informações sobre Mortalidade (SIM). Por exemplo, uma busca no site Scielo (<http://www.scielo.br/>), que reúne a maior parte da produção científica brasileira, usando combinações das palavras-chave escolaridade, educação ou instrução com mortalidade, fornece várias dezenas de artigos sobre mortalidade infantil, outras dezenas sobre mortalidade materna, e alguns outros sobre escolaridade e mortalidade por outros tipos de doenças, particularmente neoplasias. Como esta literatura não tem qualquer relação com o tema deste texto, não será discutida aqui – vale sua menção apenas para mostrar que, a despeito das dificuldades de se trabalhar com o SIM, é grande a experiência acumulada no uso da variável educação nesta base.

Há também uma literatura, um pouco menos extensa, mas também majoritariamente oriunda da área de saúde pública, sobre mortalidade violenta. Esta, no entanto, enfatiza principalmente a idade como variável correlacionada com homicídio. Os textos de Barros, Ximenes e Lima (2001) e de Batista (2005), entre outros, cobrem com bastante detalhe os diversos tipos de morte violenta, com especial ênfase sobre idade e sexo da vítima, assim como sobre a região onde ocorre a morte. Os trabalhos de Waiselfisz (1998, 2000 e 2002) e Unicef (1995) têm origem na literatura de análise educacional, mas usam abordagens semelhantes aos trabalhos oriundos da saúde pública e epidemiologia. São todos de excelente qualidade, particularmente meticolosos no tratamento dos dados do SIM. Sua principal conclusão é que a morte por homicídio está cada vez mais freqüente e mais concentrada em homens jovens. Ausente de todos, no entanto, está a variável escolaridade.

Quais trabalhos incluem a escolaridade ou a freqüência à escola como variável explicativa da taxa de homicídios no Brasil?

Carvalho, Cerqueira, e Lobão (2005) trabalham com municípios como unidade de análise, e a taxa de mortalidade por homicídio como variável a ser explicada. Os autores encontram elasticidades de 0,069 com relação à percentagem de crianças fora da escola no modelo sem autocorrelação espacial, e de 0,042 no modelo com autocorrelação espacial. Não são efeitos grandes, embora estatisticamente significativos. Na mesma linha, Gutierrez *et al.* (2004) trabalham com um painel de estados brasileiros visando investigar a relação entre desigualdade e homicídio. Os autores incluem também a freqüência à escola como variável de controle e não encontram efeitos significativos sobre a taxa de homicídios. Oliveira (2005), trabalhando com um painel de municípios e enfocando o papel do tamanho do município, encontra efeitos significativos de freqüência à escola apenas para o terceiro grau. É relevante que a variável usada em todos estes trabalhos seja a freqüência à escola, e não a escolaridade da população de quinze anos ou mais.

Néri (2006) não estuda homicídio, e sim população carcerária. No Brasil, ainda mais que nos países anglófonos, esta pode ser uma amostra muito enviesada da população perpetradora de crimes violentos, uma vez que a probabilidade de ser preso, indiciado e condenado – e o rigor da pena – pode ser função de cor ou status socioeconômico. Em todo caso, Néri mostra que a população carcerária tanto no Rio

---

6. Tradução livre.

como em São Paulo conta com nível médio de escolaridade bem menor que a população fora dos presídios.

O trabalho de Kilsztajn *et al.* (2000) tem como foco principal o papel da cor na mortalidade por homicídio em São Paulo, e trabalha com tabelas de contingência bivariadas e regressão logística multivariada. Embora o foco principal seja a cor, o estudo inclui a variável escolaridade, e esta é, de longe, a de maior impacto, apesar dos autores ignorarem quase completamente este resultado.

Soares Filho *et al.* (2007) faz em uma análise geral da mortalidade por homicídio no Brasil. Tal como Kilsztajn *et al.* (2000), Soares Filho e seus co-autores enfatizam idade, sexo e cor da pele, usando escolaridade como controle para mostrar que pretos e pardos estão mais expostos ao homicídio mesmo quando se faz o corte por nível socioeconômico. Isto de fato é demonstrado, mas novamente há um resultado muito forte que passa incólume pela análise dos autores – quatro anos ou mais de escolaridade reduzem a probabilidade de morte violenta em até 150/cem mil.

Macedo *et al.* (2001) investigam o papel dos capitais econômico e cultural na determinação das taxas de homicídio em Salvador. Em uma abordagem simples e criativa, comparam taxas de morte por homicídio de grupos com capital cultural alto e baixo e capital econômico alto e baixo. Definem capital econômico como renda e capital cultural como nível de instrução. A conclusão é que este é muito mais importante que aquele como determinante da taxa de homicídio. Enquanto o grupo capital cultural médio capital econômico baixo sofre de 14 mortes por cem mil, o grupo capital cultural baixo capital econômico baixo sofre com quase 35 – mais que o dobro. Ou seja, variando a escolaridade, mantendo a renda constante, tem-se um resultado brutal em termos de homicídio. *Mutatis mutandi*, o grupo capital cultural médio capital econômico alto sofre de quase 13 mortos por cem mil. Ou seja, variando a renda, mantendo a escolaridade constante, tem-se variações muito menores na probabilidade de morrer por homicídio.

## 2.1 RESUMO E DISCUSSÃO

A maior parte da literatura estadunidense e britânica resenhada não tem como objetivo central analisar o impacto da escolaridade sobre a propensão a cometer um crime. São trabalhos heterogêneos que variam entre os que tratam as variáveis renda e escolaridade de modo muito cuidadoso, como os de Grogger (1998) e de Lochner e Moretti (2001), até os que as incluem de modo tosco em um trabalho cujo objetivo é totalmente diverso, como os de Witte e Tauchen (1994) ou Mocan e Rees (1999). As definições do que constitui um crime também são as mais diversas, indo de prostituição, tráfico de drogas ou outros crimes sem vítima até o homicídio. A despeito destas limitações, os resultados quase sempre indicam que a escolaridade reduz a propensão ao crime. Em trabalhos feitos com este objetivo específico, Harlow (2003) e Lochner e Moretti (2001) encontram resultados fortes e significativos para o término do segundo grau.

No Brasil, *grosso modo*, há duas linhas de pesquisa: uma oriunda da literatura econômica e outra da literatura da saúde pública. Ambas tratam apenas do crime de homicídio, e o fazem do ponto de vista das vítimas, uma vez que este se encontra bem documentado nos dados da saúde pública fornecidos pelo SIM. O ramo da literatura

econômica usa painéis de estados, regiões metropolitanas ou municípios, não tem como objetivo a investigação específica da relação entre educação e homicídio, e encontra majoritariamente resultados fracos, embora geralmente significativos.

Já o ramo da literatura de saúde pública usa principalmente recortes por características individuais como idade, sexo e cor da pele, também não tem como objetivo específico a análise do impacto da educação, e encontra resultados muito fortes. Acreditamos que esta abordagem é mais promissora que a abordagem territorial usada pelos economistas. É curioso que os autores da saúde pública não tenham dado muita importância aos impactos muito fortes que encontraram do nível de escolaridade sobre a probabilidade de se morrer vítima de homicídio. Imagina-se que isto tenha ocorrido porque os autores não estavam à procura deste resultado.

O trabalho de Macedo *et al.* (2001) é o único que tem como foco principal o efeito da escolaridade sobre a probabilidade de se morrer vítima de homicídio. No entanto, ele cobre apenas Salvador e usa agrupamentos pouco finos de renda e escolaridade. O objetivo deste estudo é aprofundar a análise de Macedo *et al.* (2001), usando ferramentas como as de Kilsztajn *et al.* (2000) e Soares Filho *et al.* (2007), e buscando fazer inferência para todo o território nacional.

A hipótese que embasa a construção do modelo econométrico, cujos detalhes se encontram na próxima seção e que o leitor sem interesse matemático pode abstrair sem prejuízo à compreensão do restante do texto, é que a educação protege o indivíduo de ser vítima de homicídio. Mas como se daria tal proteção?

Imaginamos aqui que qualquer descrição adequada do processo que leva a um homicídio envolve várias relações. Uma descrição simples inclui no mínimo duas: o grau de vulnerabilidade da vítima e a propensão de ela estar em situações ou ambientes nos quais o conflito letal é provável. Uma viúva idosa, por exemplo, dificilmente estará em ambientes ou situações de possível conflito letal, mas é altamente vulnerável, o que possivelmente explica o aumento na taxa de morte por homicídio de mulheres com idades acima de 50 anos, visível no gráfico 3 adiante (subseção 4.1, p. 20). Já um jovem de sexo masculino talvez não seja tão vulnerável, mas está com frequência em situações de risco, que podem variar de brigas de bar até o engajamento em uma carreira criminal, onde o conflito é possível. O relevante é que uma descrição adequada do processo de vitimização por homicídio envolve duas ou mais equações diferentes.

Infelizmente, os dados disponíveis impõem limites severos. Não é possível observar com precisão estatística situações ou ambientes nos quais o conflito letal seja provável. Estes limites nos obrigam à estimação de uma forma reduzida de uma única equação do modelo estrutural, o que significa dizer que a interpretação dos coeficientes deve ser feita sempre com muito cuidado.

No entanto, ainda que não seja estimável, a admissão de que a descrição estatística ideal do homicídio é um modelo estrutural joga luz sobre a ambigüidade que acompanha este texto do início até o seu final: em que medida há uma identidade estatística entre vítimas e perpetradores? A resposta intuitiva é que em uma ou mais das equações há esta identidade, e em outras, não. Por exemplo, deve haver considerável identidade entre vítimas e perpetradores na equação que define exposição a situações de risco: homens pobres jovens negros vivendo em áreas

metropolitanas são um grupo cujas vidas estão quase continuamente sujeitas a situações de risco. Em termos de vulnerabilidade, contudo, é possível que esta identidade não exista. Infelizmente, a forma reduzida não permite observar estas diferenças e, portanto, gera problemas na interpretação dos coeficientes, e esta identidade imperfeita será sempre uma dificuldade.

A estimação da forma reduzida leva a outras dificuldades. Por exemplo, a interpretação do impacto da renda torna-se também difícil. Um aumento de renda terá um efeito não ambíguo sobre a exposição a situações de risco: um indivíduo rico viverá, se transportará e trabalhará em áreas mais policiadas e mais seguras. Entretanto, um indivíduo de renda alta será um alvo mais tentador para todo homicídio relacionado com ganho pecuniário.

Expostas estas dificuldades, a próxima seção lida com os problemas econométricos propriamente ditos.

### **3 DADOS E METODOLOGIA**

A variável dependente deste trabalho é a taxa de homicídios por 100 mil habitantes. Sendo a taxa de homicídio uma taxa, é preciso saber tanto quantas pessoas morreram como quantas pessoas havia no total em uma dada população em um dado ano. Infelizmente, nenhuma fonte de dados no Brasil contém ambas as informações, e por isso temos que buscar o numerador e o denominador em bases de dados diferentes. Esta dificuldade levará a toda uma série de dificuldades metodológicas no que concerne a pareamento de observações, estimação de população total e comportamento de não-resposta.

#### **3.1 FONTES DE DADOS**

A fonte de dados para o numerador é o Sistema de Informações sobre Mortalidade (SIM) entre 1999 a 2004. O SIM é um registro administrativo coletado anualmente pelo Sistema Único de Saúde (SUS), e é razoavelmente completo no que tange ao registro de todas ou quase todas as mortes ocorridas no Brasil, inclusive as mortes violentas. Registram-se também no SIM algumas variáveis que podem estar associadas à probabilidade de morte, tais como idade, data de nascimento, sexo, raça/cor, estado civil e escolaridade, além de variáveis geográficas explicitando onde e quando ocorreu a morte. Estas variáveis adicionais nem sempre estão bem coletadas, e o grau de precisão e de dados ausentes depende muito da Unidade da Federação (UF) e do ano em questão. Há um bom resumo do SIM e suas dificuldades nos trabalhos de Waiselfisz (1998, 2000, 2002 e 2004), além de no próprio site do SIM (<http://www.datasus.gov.br/catalogo/sim.htm>).

Os códigos que se referem a homicídio neste trabalho são oriundos da 10ª Classificação Internacional de Doenças (CID 10): agressões (X85-Y09) e intervenções legais e operações de guerra (Y35-Y36).

Já as fontes de dados para o denominador são o Censo Demográfico de 2000 e as Pesquisas Nacionais por Amostra de Domicílios do IBGE (Pnads/IBGE) de 1999 a 2004. Estas são as únicas duas fontes de dados que cobrem todo (ou quase todo) o território nacional e contêm a população total, além de dados sobre idade, data de

nascimento, sexo, raça/cor, estado civil e escolaridade. O uso do censo não tem mistérios. O número de observações da amostra ponderados pelos pesos<sup>7</sup> fornecidos pelo IBGE dá o resultado da população para qualquer corte que possa ser feito no SIM, permitindo assim o cálculo de taxas de homicídio sem qualquer outro ajuste, a não ser pareamento e comportamento de não-resposta, os quais serão vistos a seguir.

As Pnads apresentam erro amostral que invariavelmente irá introduzir viés nas estimações. No entanto, como não desejamos aqui ficar circunscritos apenas ao ano 2000, é necessário usar fontes que são coletadas anualmente, como a Pnad, ou então projeções populacionais independentes. O uso de projeções populacionais independentes nos limitaria ao uso das variáveis incluídas nestas projeções, normalmente apenas idade e sexo. As Pnads contêm uma riqueza de informações, ausentes tanto no SIM como nas projeções populacionais, que poderiam contribuir para analisar o fenômeno. O uso destas pesquisas leva a uma outra dificuldade, que é a correção das respectivas projeções populacionais.

### 3.2 METODOLOGIA

Seguindo Kilsztajn *et al.* (2005) e os outros estudos da tradição da epidemiologia, usaremos neste texto tanto tabelas de contingência bi e trivariadas como análise de regressão logística com dados agrupados. No caso da análise das tabelas de contingência, serão usadas as taxas de homicídio por 100 mil habitantes.

A análise de regressão é mais complicada, uma vez que os dados aqui utilizados advêm de duas fontes: o SIM para mortes e o censo demográfico ou a Pnad para população total. Como nem uma nem outra identificam indivíduos, será necessário identificá-los mediante o pertencimento a grupos definidos por variáveis em comum. Ou seja, será necessário trabalhar com dados agrupados.

Segundo Maddala (1983), o modelo linear de probabilidade com dados agrupados não se encontra sujeito a críticas tão fortes como com dados individuais. De fato, se em todas as células a probabilidade de ocorrência empírica se situa no intervalo (0,1), os coeficientes podem ser estimados mediante a seguinte regressão:

(1)  $\hat{p}_i = \beta x_i + u_i$  onde  $x_i$  representa um vetor de variáveis explicativas,  $\beta$  os coeficientes estimados,  $u_i$  um erro idiossincrático e  $\hat{p}_i$  as probabilidades observadas. A estimação é feita por mínimos quadrados ponderados usando como pesos  $w_i$  onde:  $w_i = \sqrt{\frac{n_i}{\hat{p}_i(1-\hat{p}_i)}}$ , e onde  $n_i$  é o número de observações na célula  $i$ .

Dado que o homicídio é um evento raro em termos estatísticos, é possível que o modelo linear de probabilidade não seja muito preciso por não oferecer uma forte resolução para eventos cuja probabilidade é muito próxima de zero. Embora o uso de dados agrupados reduza muito este problema, convém lembrar que uma das críticas ao modelo linear é que permite probabilidades inferiores a zero e superiores a um, uma vez que a propensão latente,  $X\beta$ , se converte diretamente em probabilidade. Isto não é um problema para eventos cuja frequência encontra-se longe dos extremos, mas é um problema quando aproximamo-nos das fronteiras naturais da probabilidade:

---

7 O censo demográfico também é uma amostra, apesar de ser uma amostra muito grande.

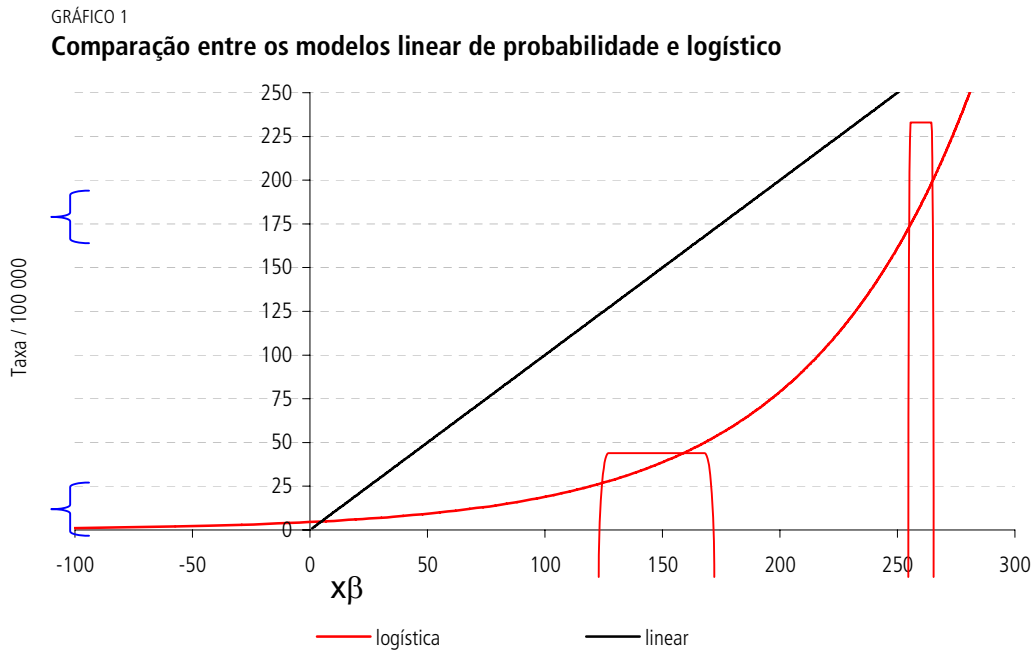


zero e um. Esta é uma das razões que levam pesquisadores a modelarem eventos usando uma função de probabilidade para transformar a propensão latente em probabilidade:  $P = F(X\beta)$ . Uma das mais comuns é a função logística, que leva ao modelo de regressão logística.

Novamente segundo Maddala (1983), é possível estimar uma regressão logística com dados agrupados mediante o uso de mínimos quadrados ponderados:

$$(2) \quad \ln \frac{\hat{p}_i}{1 - \hat{p}_i} = \beta x_i + u_i \quad \text{onde os pesos são: } w_i = \sqrt{\frac{1}{n_i \hat{p}_i (1 - \hat{p}_i)}}$$

O gráfico 1 compara as funções linear de probabilidade e logística. Na distribuição hipotética logística a seguir,<sup>8</sup> se o impacto de uma dada variável sobre a probabilidade de homicídio é elevar a taxa de 25 para 50 por 100 mil, o produto desta variável com seu coeficiente,  $X\beta$ , é de aproximadamente 50 contra 25 da função linear. Já se o aumento for os mesmos 25 por 100 mil, mas indo de 175 para 200, a função logística terá uma variação de 10 contra os mesmos 25 da função linear.



Elaboração do autor.

Ou seja, apesar de no caso do homicídio as probabilidades estarem próximas de zero, a grande variação entre taxas de diferentes grupos impede que se eleja, *a priori*, um modelo enquanto superior a outro. Uma vantagem adicional do modelo linear é ser altamente confiável. Mesmo quando várias das hipóteses não são verificadas, ainda assim os mínimos quadrados ordinários produzem estimativas consistentes.

Note-se que nem o modelo linear nem o modelo logístico admitem grupos de dados nos quais a probabilidade de ocorrência seja zero ou unitária. Isto impõe um limite ao

8. O exemplo foi escolhido para ilustrar probabilidades da ordem de grandeza da morte por homicídio. Outros parâmetros trarão outros números, mas o comportamento qualitativo será o mesmo, com a logística oferecendo resolução maior para eventos mais raros, e a linear, resolução maior para eventos mais frequentes.

nível de desagregação máximo aceitável – tem que haver pelo menos um homicídio em todas as células – e será um dos muitos problemas detectados ao longo deste texto.

### 3.2.1 O pareamento usando células

O modelo de dados agregados pode ser aplicado sem grandes dificuldades para o ano 2000, para o qual dispomos de um censo demográfico e dos dados, também censitários, do SIM. Para este ano, agrupamos os dados de ambas as bases em células definidas por idade – dividida em 17 grupos –, sexo, cor – dividida em negros e não-negros, com indígenas incluídos na categoria negros e amarelos incluídos na categoria não-negros,<sup>9</sup> nível educacional – dividido em seis grupos –, a grande região de residência, e, finalmente, se o indivíduo mora ou não em região metropolitana. Isto perfaz um total de 2.040 células.

Não houve problema do ponto de vista do censo, que forneceu totais populacionais adequados, mas em quase metade das células não houve mortes por homicídio, o que invalidaria a metodologia usando dados agregados. Em função disso, optamos por reduzir o número de células mediante quatro agregações. A primeira foi reduzir o número de categorias de idade para o sexo feminino, nas quais há maior número de células sem ocorrência de homicídio. A segunda foi considerar apenas dois aglomerados de regiões: Norte e Nordeste, por um lado, e Sudeste, Sul e Centro-Oeste, por outro. A terceira foi eliminar da amostra todos os indivíduos com menos de 14 anos. A quarta foi suprimir o uso da variável região metropolitana. Finalmente, eliminamos as células correspondentes a indivíduos com idade entre 14 e 20 anos sem instrução ou com instrução superior. Estas restrições resultaram em 288 células e reduziram para apenas dois (menos de 1%) o número de células sem homicídios. Dado que o limitador é o número de homicídios do SIM, o mesmo procedimento foi adequado também para o pareamento com a Pnad.<sup>10</sup>

Posto que o Censo Demográfico de 2000 fornece dados relativos apenas para aquele ano, será necessário apelar para as Pnads, que são pesquisas domiciliares amostrais anuais. O esquema amostral das Pnads tem como único objetivo estimar os totais populacionais de certos estratos amostrais geográficos (*UF x tipo de município*) e não oferecem totais populacionais precisos por idade e sexo.<sup>11</sup> Tendo em vista que vamos parear informações da Pnad com informações do SIM, é fundamental que estes totais sejam os mais exatos possível. Portanto, há necessidade de se padronizarem os pesos da Pnad para que estes reflitam o tamanho das coortes por idade simples na população total.

A Pnad também tem a desvantagem de, até 2004, não incluir a área rural da região Norte. Já no SIM, é impossível identificar a área de residência (rural ou não-

---

9. Ainda assim, esta deve ser uma das variáveis mais problemáticas deste estudo, uma vez que tanto o censo como a Pnad se valem da auto-identificação – ou seja, é o próprio indivíduo que se declara ou não como negro – e, por razões óbvias, no SIM, o procedimento usado é a hetero-identificação – ou seja, é o médico legista quem classifica o indivíduo como negro ou não. Isto é particularmente grave quando se considera que freqüentemente um corpo no IML se encontra despido não apenas de suas roupas, mas também de sinais que identifiquem o pertencimento a uma ou outra categoria racial.

10. A exclusão da região Norte não muda o número de células, uma vez que as regiões Norte, Centro-Oeste e Sudeste são agrupadas para fins de pareamento.

11. É claro que é possível calcular totais por raça ou sexo. O que se quer destacar aqui é que estes não são estratos amostrais da Pnad e, portanto, qualquer total calculado estará sujeito a um erro amostral. O procedimento explicado adiante visa dirimir este erro.

rural) da vítima.<sup>12</sup> Desse modo, para tornar os dados da Pnad e do SIM compatíveis, adotamos dois procedimentos para compatibilizar a abrangência regional e a população por faixa etária.

Primeiro, eliminamos no SIM, na Pnad e no censo demográfico todos os registros referentes aos estados do Acre, Amapá, Pará Amazonas, Roraima e Rondônia. Mantivemos, da região Norte, apenas o estado de Tocantins, cuja área rural é incluída na Pnad, uma vez que seu desenho foi feito quando este ainda fazia parte do estado de Goiás.

Em segundo lugar, usamos a estrutura etária por sexo presente no censo demográfico (sem a região Norte, conforme explicado) e aplicamos à mesma as probabilidades de morte por sexo e idade simples fornecidas nas tábuas de vida do IBGE (<http://www.ibge.gov.br/home/estatistica/populacao/tabuadevida/>). Dado que a estrutura etária no censo diz respeito à idade quando da realização da entrevista, e não ao ano explícito de nascimento, e ainda o fato de o censo ser realizado ao longo de todo o ano de 2000, utilizamo-nos da média das probabilidades de morte do ano base  $t$  e o ano  $t + 1$  para estimar a população por idade e sexo em  $t + 1$ . Fazendo isso usando como base o ano 2000, chegamos à população em 2001; aplicando a mesma metodologia de modo iterado, chegamos à população em qualquer ano  $t + n$ , desde que existam tábuas de vida para todos os anos intermediários. De modo análogo, é possível retroceder no tempo para chegar às estruturas etárias de anos anteriores a 2000.

Ou seja, os pesos usados na construção das células usando as Pnads são os seguintes:

$$(3) \quad P_{j,i,k}(A) = Q_{j,i,k}(A) \left( \frac{Pop_{censo}(i,k) \prod_{a=2000}^{2000+A} \left[ 1 - \left( \frac{t(i,k,a) + t(i,k,a+1)}{2} \right) \right]}{Pop_{pnad}(i,k,A)} \right)$$

onde  $P_{j,i,k}(A)$  é o peso do indivíduo  $j$  do sexo  $k$  nascido no ano  $i$  a ser calculado no ano  $A$ ,  $Q_{j,i,k}(A)$  é o peso fornecido pelo IBGE para o mesmo indivíduo na Pnad,  $Pop_{censo}(i,k)$  é a população do sexo  $k$  nascida no ano  $i$  medida no Censo de 2000,  $Pop_{pnad}(i,k,A)$  é a população nascida em  $i$  de sexo  $k$  medida na Pnad do ano  $A$ , e  $t(i,k,a)$  é a taxa de mortalidade por todas as causas da população nascida no ano  $i$  com sexo  $k$  e observada no ano  $a$ , com  $a$  variando de 2000 até  $A$ .

Anos anteriores a 2000 são calculados com uma equação análoga, apenas trocando o termo no produto por  $\left[ 1 + \left( \frac{t(i,k,a) + t(i,k,a+1)}{2} \right) \right]$ , uma vez que cada coorte fica maior – e não menor – quando se retrocede no tempo.

Esta metodologia faz três hipóteses implícitas fortes. A primeira é que as probabilidades de morte por sexo e idade simples são semelhantes para o Brasil como um todo e para o Brasil fora região Norte. A segunda é que as tábuas de vida são suficientemente precisas para serem usadas de modo iterativo, pelo menos para quatro

12. Seria possível fazer uma tipologia de municípios enquanto predominantemente rurais ou predominantemente urbanos, mas estes não seriam identificáveis na Pnad – apenas no censo. Por isso optamos por usar médias por célula também nesta variável.

anos. Finalmente, há a hipótese de que não há saldo migratório significativo do Brasil – fora região Norte – para a região Norte ou para fora do país. É claro que nenhuma destas hipóteses se verifica de modo exato, o que sugere que os resultados para 2001 serão mais precisos que os resultados para, digamos, 2005.

### 3.2.2 Comportamento de não-resposta

Um problema grave para o pareamento do SIM, seja com o censo, seja com a Pnad, se refere à taxa diferencial de não-resposta em algumas variáveis importantes. As taxas de não-resposta tanto no censo como na Pnad são muito baixas: para cor, idade e sexo quase não há não-resposta e as taxas são todas inferiores a 0,01%. Não-resposta é um pouco maior no caso de educação, cuja taxa chega a 0,8% para pessoas com 10 anos ou mais. Já o SIM é preenchido por legistas que freqüentemente não têm outra informação além do corpo da vítima e há taxas altas de não-resposta: idade (0,46%), cor (9%) e até sexo (0,1%). Para nível de instrução, as taxas chegam a 37%.

Optamos aqui por usar dois tratamentos diferentes para não-resposta no SIM. O primeiro foi eliminar indivíduos com idade desconhecida tanto da Pnad como do SIM. O segundo seguiu uma metodologia chamada *hot deck imputation* e foi usado para as variáveis cor, sexo e escolaridade.

O *hot deck imputation* consiste em dividir os dados em estratos o mais homogêneos possíveis definidos por variáveis sem não-resposta, e usar o dado informado pelo vizinho mais próximo como doador de informação para o caso não informado. Os estratos que usamos para tanto foram formados pelas variáveis: *i)* ser ou não vítima de homicídio; *ii)* a UF onde ocorreu e morte; e *iii)* o município onde ocorreu a morte. No caso da escolaridade, a idade também foi considerada um estrato. Em termos práticos, o *hot deck* funciona assim:

- 1) As observações foram agrupadas por UF onde a morte ocorreu, e, dentro de cada UF, foram agrupadas por município.<sup>13</sup>
- 2) Dentro de cada município, houve ainda um segundo agrupamento. Para sexo e raça, as observações foram ordenadas pela data de ocorrência. Para nível de escolaridade, foram primeiro ordenadas por idade e depois por data de ocorrência.
- 3) Com os dados assim agrupados, para cada observação sem resposta, procurou-se primeiro a observação posterior. Se esta tivesse um valor válido para a variável, este valor foi doado para a observação sem resposta. Caso contrário, procurou-se a seguir a observação anterior, duas antes, duas depois, três antes, três depois e assim por diante.

A técnica é simples e tem boas propriedades mesmo na presença de uma variedade de comportamentos de não-resposta, conforme mostrado por Hu, Salvucci e Cohen (1998). Seastrom, Kaufmann e Lee (2003) contém uma boa descrição e algoritmos para fazer *hot deck imputation*.

---

13. Há uma aparente redundância ao utilizar UF e município. Utilizamos os dois para proteger as estimativas de eventuais erros de preenchimento de código do município. Na pior das hipóteses, utilizamos apenas um pouco mais de tempo de computador.

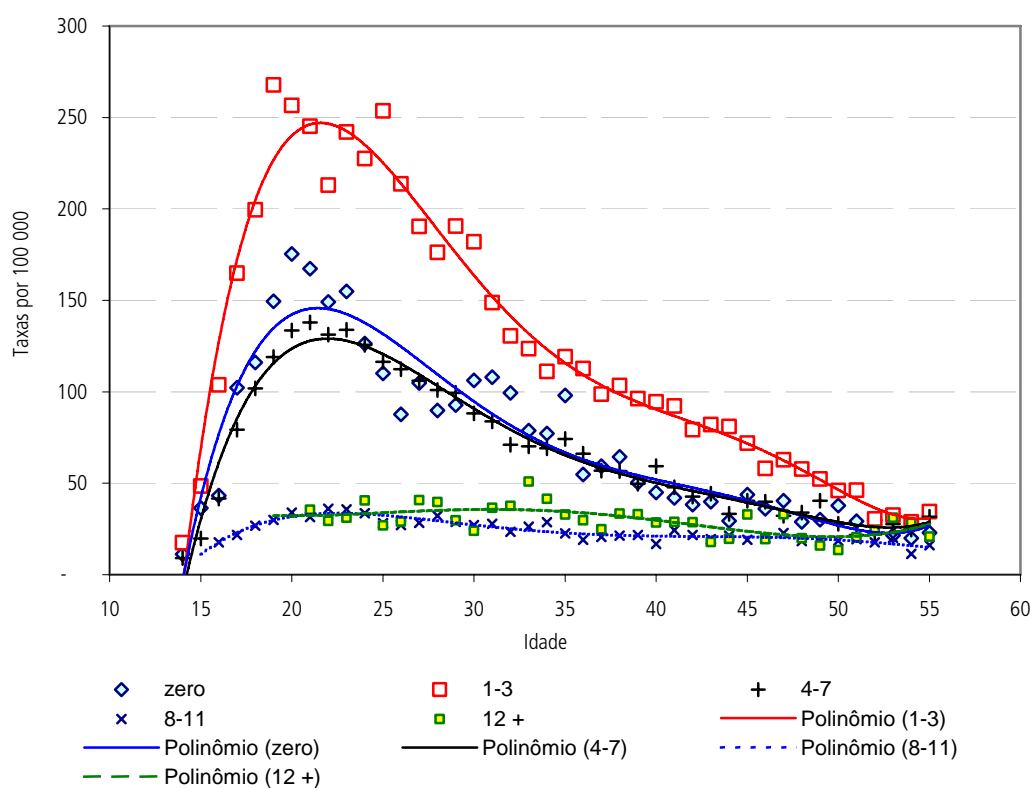
## 4 RESULTADOS

### 4.1 TABELAS DE CONTINGÊNCIA

O primeiro resultado apresentado é justamente o que motivou a elaboração deste texto. Trata-se do resultado mostrado no gráfico 2, que mostra as taxas de homicídios por 100 mil pessoas por idade e nível de escolaridade. Cada curva corresponde a um nível de escolaridade de acordo com a subdivisão disponível no SIM (nenhum estudo, menos que o primeiro ciclo do ensino fundamental, entre o primeiro e o segundo ciclo do ensino fundamental, entre fundamental completo e médio completo, e algum ensino superior ou mais).

GRÁFICO 2

#### Taxas de homicídios por idade simples e escolaridade – Homens



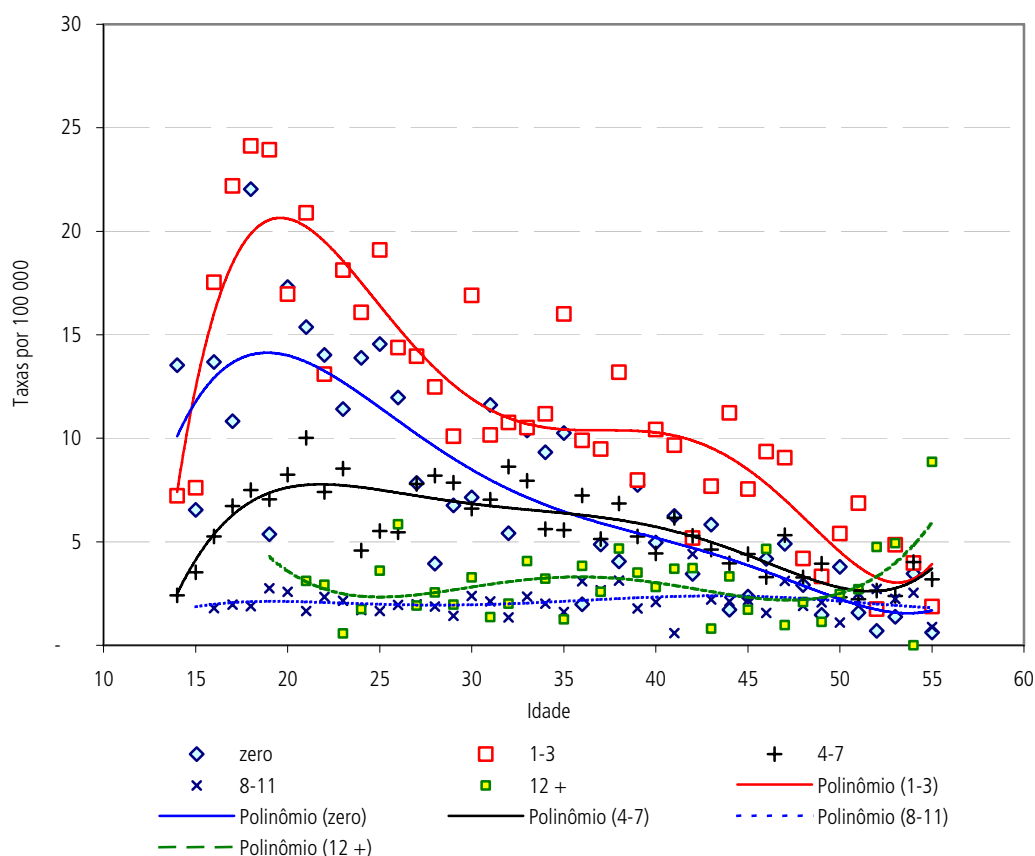
Fontes: SIM (2000) e Censo Demográfico (2000).

A forma das curvas – muito altas para idades entre 16 e 36 anos – é fato conhecido há muito tempo na literatura sobre homicídio. Note-se que as curvas seguem aproximadamente o mesmo padrão, com máximo próximo de 21 anos e assimetricamente concentradas entre os mais jovens que 21 anos. O que surpreende em especial é a distância entre as curvas. A altura do pico da curva de morte por homicídios dos jovens com 1 a 3 anos de escolaridade corresponde ao dobro da altura da curva de morte dos que terminaram pelo menos o primeiro ciclo do fundamental, e a quase oito vezes a altura do pico para os que têm entre o fundamental completo e o médio completo. É ainda mais interessante que a diferença entre as curvas destes últimos e a dos que frequentaram algum curso superior não é tão grande.

O gráfico 3 mostra as mesmas curvas para mulheres. Apesar do formato da curva ser levemente diferente – o pico ocorre um ou dois anos antes e a queda após o pico da taxa de homicídio é um pouco mais lenta –, o principal resultado é que o efeito da escolaridade parecer ser mais ou menos o mesmo, em termos multiplicativos, que entre os homens. Ou seja, apesar das taxas de homicídio femininas serem aproximadamente um décimo das taxas masculinas, a diferença entre os níveis de escolaridade é mais ou menos equivalente: o dobro entre 1 a 3 e 4 a 7 anos; um fator de quase oito entre 1 a 3 e 8 a 11 anos; e quase o mesmo entre 8 a 11 anos e 12 anos ou mais.

GRÁFICO 3

**Taxas de homicídios por idade simples e escolaridade – Mulheres**



Fontes: SIM (2000) e Censo Demográfico (2000).

Uma das primeiras explicações que vêm à mente é que pelo menos parte desse efeito não é um efeito direto do percurso escolar de cada indivíduo, e sim um efeito renda. Afinal, é fato notório que a escolaridade é um dos maiores determinantes da renda no Brasil – e em quase todos os países do mundo.

A tabela 1 sugere que não é assim, uma vez que mostra não serem os aumentos na renda domiciliar *per capita* grandes entre as primeiras três categorias de escolaridade disponíveis no SIM, mas muito grandes entre estas e as duas últimas. Em particular, chama atenção que há um grande aumento de renda entre a categoria 8-11 e 12 ou mais, cujos níveis de homicídio são mais ou menos equivalentes.

No entanto, a tabela 1 pode apenas sugerir. A investigação mais detalhada da relação entre escolaridade, homicídio e renda requer análise de regressão múltipla

para isolar o efeito do nível escolar de um eventual efeito da renda sobre a probabilidade de ser morto por outro indivíduo.

TABELA 1

**Taxas de homicídios por grupos de idade, escolaridade e sexo**

		Nível de escolaridade					Total	
		Homens	zero	1-3	4-7	8-11	12 +	
Faixa etária	14-17		48,33	74,53	31,93	17,59	-	35,96
	18-25		142,66	237,67	124,23	31,76	44,42	95,31
	26-35		94,30	159,01	88,31	27,32	36,19	72,94
	36-55		38,07	70,99	42,68	19,97	25,66	39,15
		Mulheres						
Faixa etária	14-17		11,03	12,48	3,96	1,92	-	4,52
	18-25		14,23	18,97	7,36	2,12	3,09	5,79
	26-35		8,46	12,67	7,08	1,90	2,74	5,38
	36-55		3,22	7,28	4,57	2,34	3,07	4,09
Renda domiciliar <i>per capita</i> (14 a 55)		R\$ 112	R\$ 138	R\$ 193	R\$ 364	R\$ 1.072	R\$ 320	

Fontes: SIM (2000), Censo Demográfico (2000) e Pnad (2001).

## 4.2 REGRESSÃO MULTIVARIADA

Conforme explicamos na metodologia, a impossibilidade de parear dados individuais entre o SIM, por um lado, e censo ou Pnad, por outro, obrigou o autor a usar dados agrupados em células definidas por sexo, grupos de idade, cor, região e, é claro, nível de escolaridade. Começamos a análise com o ano de 2000, para o qual o censo forneceu o número total de pessoas em cada célula, e o SIM, o número de homicídios, fazendo o cálculo da taxa de homicídio da célula tarefa simples.

Além das próprias variáveis que definem a célula, calculamos as médias, por célula, de duas outras variáveis – a renda domiciliar *per capita* e a taxa de urbanização.

Ainda de acordo com o que já foi aqui colocado, as regressões foram estimadas mediante o uso de dois métodos de estimação. O primeiro foi o modelo linear de probabilidade, que consiste na estimação direta dos determinantes da taxa de homicídio por mínimos quadrados. Este método de estimação apresenta duas vantagens grandes. Primeiro, o efeito marginal sobre as taxas de homicídio por 100 mil pode ser lido diretamente dos próprios coeficientes. Segundo, é muito robusto e fornece coeficientes consistentes mesmo quando muitas das hipóteses básicas são violadas.

O segundo método foi uma regressão logística, que consiste na estimação do logaritmo da razão entre as probabilidades de morrer e não morrer vítima de homicídio. A principal vantagem da regressão logística é que a função logística aumenta a resolução e precisão próximas do ponto zero (ou um), o que gera estimações de melhor qualidade quando se analisam eventos raros – e o homicídio é, felizmente, até no Brasil, um evento raro do ponto de vista estatístico. As estimações foram feitas com a metodologia envolvendo mínimos quadrados ponderados descrita na subseção sobre metodologia (subseção 3.2).

Para cada método, foram estimados três modelos. O primeiro envolve apenas as variáveis encontradas nos gráficos 2 e 3 e está presente como verificação de que o modelo traduz corretamente a análise tabular mostrada nos gráficos. No segundo, levamos em conta todas as variáveis que definem células mediante a inclusão de cor e região (agrupadas apenas em duas grandes regiões). Por fim, no terceiro modelo

inserimos as médias por célula da renda domiciliar *per capita* e a taxa de urbanização. *A priori*, o terceiro modelo é o mais confiável por incluir a variável renda.

TABELA 2  
Resultados de regressão para 2000

	Modelo linear de probabilidade						Regressão logística					
	Homens		Mulheres		Rural		Homens		Mulheres		Rural	
	$\beta$	p-valor	$\beta$	p-valor	$\beta$	p-valor	$\beta$	p-valor	$\beta$	p-valor	$\beta$	p-valor
<b>Idade</b>												
Idade 10	0,0		0,0		0,0							
Idade 15	53,9	0%	57,3	0%	68,4	0%	3,158	0%	3,237	0%	3,031	0%
Idade 18	93,2	0%	95,8	0%	110,2	0%	4,215	0%	4,266	0%	3,993	0%
Idade 21	110,9	0%	112,8	0%	130,0	0%	4,506	0%	4,399	0%	4,080	0%
Idade 24	105,9	0%	107,7	0%	123,7	0%	4,393	0%	4,400	0%	4,122	0%
Idade 27	96,4	0%	98,1	0%	112,6	0%	3,723	0%	3,846	0%	3,608	0%
Idade 30	93,2	0%	94,9	0%	106,6	0%	3,917	0%	3,996	0%	3,786	0%
Idade 33	89,2	0%	87,9	0%	99,3	0%	4,459	0%	4,424	0%	4,213	0%
Idade 36	87,5	0%	86,3	0%	97,6	0%	4,387	0%	4,351	0%	4,136	0%
Idade 40	72,7	0%	71,0	0%	79,7	0%	4,109	0%	4,066	0%	3,893	0%
Idade 50	45,7	0%	43,4	0%	40,6	0%	3,691	0%	3,614	0%	3,567	0%
<b>Escolaridade</b>												
zero	0,0		0,0		0,0							
1-3	18,2	15%	20,4	8%	12,9	30%	-0,256	33%	-0,023	88%	0,031	85%
4-7	-1,4	91%	-2,7	80%	-23,9	14%	-0,646	1%	-0,372	2%	0,016	94%
8-11	-53,4	0%	-52,8	0%	-99,0	0%	-2,278	0%	-1,932	0%	-1,123	0%
12 +	-39,4	1%	-39,6	0%	-151,2	0%	-1,212	0%	-1,059	0%	0,215	68%
<b>Cor/raça</b>												
negra			19,5	0%	28,5	0%			1,077	0%	0,969	0%
branca			0,0		0,0							
<b>Região</b>												
Nordeste, Norte			-31,3	0%	-16,3	8%			-1,016	0%	-1,309	0%
Sudeste, Sul, CO			0,0		0,0							
Renda dom. <i>per capita</i> (R\$ 100)					7,5	3%					-0,001	20%
Rural					-67,9	22%					1,889	1%
Constante	-10,3	42,%	-2,0	87%	-17,9	47%	-11,23	0%	-11,13	0%	-12,71	0%
<b>Mulheres</b>												
<b>Idade</b>												
Idade 10	0,0		0,0		0,0							
Idade 20	7,3	0%	7,5	0%	8,3	0%	2,639	0%	2,734	0%	2,500	0%
Idade 30	7,2	0%	7,2	0%	7,9	0%	2,674	0%	2,725	0%	2,521	0%
Idade 40	6,4	0%	6,4	0%	6,9	0%	2,443	0%	2,522	0%	2,389	0%
Idade 50	3,7	0%	3,6	0%	3,0	1%	2,209	0%	2,218	0%	2,245	0%
<b>Escolaridade</b>												
zero	0,0		0,0		0,0							
1-3	1,9	17%	2,1	12%	1,7	24%	-0,153	62%	-0,018	94%	0,080	74%
4-7	0,2	88%	0,2	86%	-0,7	70%	-0,412	17%	-0,240	31%	0,228	46%
8-11	-3,8	1%	-3,8	0%	-6,1	7%	-1,612	0%	-1,429	0%	-0,487	33%
12 +	-2,2	18%	-2,3	14%	-8,0	17%	-0,646	1%	-0,524	1%	0,753	36%
<b>Cor/raça</b>												
negra			0,9	18%	1,5	7%			0,519	0%	0,445	1%
branca			0,0		0,0							
<b>Região</b>												
Nordeste, Norte			-2,3	0%	-1,5	26%			-0,745	0%	-1,113	0%
Sudeste, Sul, CO												
Renda domiciliar <i>per capita</i> (R\$ 100)					0,5	23%					0,000	62%
Rural			0,0		-2,0	80%					2,915	2%
Constante	-0,8	58%	-0,1	95%	-0,8	74%	-12,23	0,8	-12,14	0%	-12,73	0%

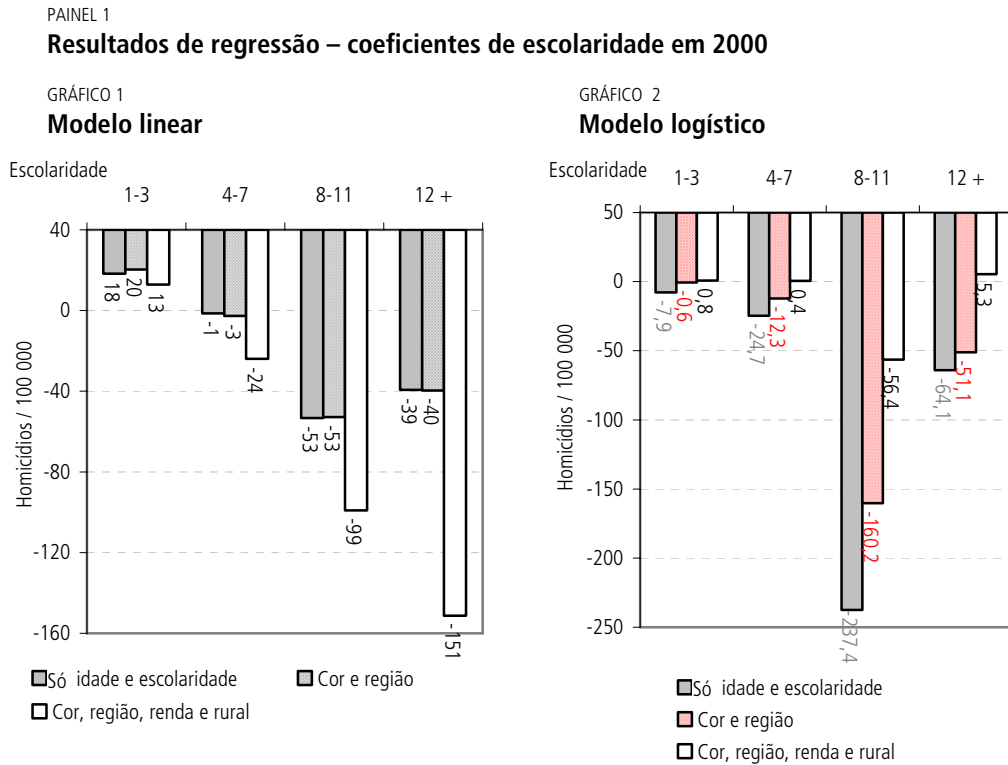
Fontes: SIM (2000) e Censo Demográfico (2000) – inclui a região Norte.

Nota: Efeitos marginais na regressão logística calculados na taxa de homicídio: média de 2000.

A tabela 2 contém muitos números, mas nem todos são igualmente importantes. Alguns deles confirmam fatos conhecidos de todos os que trabalham com o tema homicídio. Primeiro, negros morrem mais que brancos: mesmo com todos os controles, a cor negra aumenta a chance de morrer por homicídio em algo entre 20 e 30 por 100 mil para homens e próximo de um por 100 mil para mulheres. Segundo, morar no Sudeste é perigoso. A célula que inclui o Sudeste também inclui o Sul e o



Centro-Oeste, e acredita-se que, se fosse possível aqui separar as regiões, seria a região Sudeste – especialmente os estados do Rio de Janeiro, Espírito Santo e São Paulo – que elevaria a média. Finalmente, no modelo linear, a idade se comporta quase exatamente como nos gráficos 2 e 3. Todos estes resultados dependem pouco da inclusão de outras variáveis na regressão.



Os resultados para escolaridade, no entanto, mudam consideravelmente quando a variável renda é incluída. Por esta razão, incluímos um gráfico apenas destes coeficientes para homens. O modelo linear e o modelo logístico contam, no entanto, histórias distintas. O primeiro gráfico do painel 1 mostra uma história na qual a renda aumenta a probabilidade de um indivíduo morrer vítima de homicídio, enquanto a escolaridade a diminui. Em função disso, há fortes variações nos coeficientes quando se inclui a variável renda domiciliar *per capita*. A proteção oferecida por algum estudo de nível superior *vis-à-vis* 1 a 3 anos de escolaridade aumenta de 57 para 164 mortos por 100 mil habitantes quando se controla pela renda.<sup>14</sup> A proteção oferecida pelo segundo grau (categoria 8-11) *vis-à-vis* 1 a 3 anos de escolaridade aumenta de 71 por 100 mil para 112.

Já cada R\$100,00 de renda domiciliar *per capita* aumenta a chance de morrer por homicídio em 7,5 por 100 mil. Ou seja, ir do centésimo 50 para o 90 na distribuição de renda aumenta em 32 por 100 mil as chances de morrer por homicídio. Se, de fato,

14. A categoria omitida é 'zero ano de estudo'. Isto quer dizer que a diferença entre a categoria mais sujeita a risco (1 a 3 anos de escolaridade) e a menos sujeita a risco (12 ou mais) se obtém pela diferença dos respectivos coeficientes: no caso do modelo que inclui a renda,  $13 - (-151) = 13 + 151 = 164$ .

podermos interpretar os coeficientes na tabela 2 como causais, conclui-se que a escola oferece um impressionante escudo contra a morte por homicídio.

No entanto, o segundo gráfico do painel 1 conta outra história, na qual a renda oferece quase toda a proteção. A inclusão da renda domiciliar *per capita* faz com que o único nível de escolaridade cujo coeficiente é significativamente diferente de zero seja o nível médio. O coeficiente da renda, por sua vez, apesar de significativo apenas a 20%, tem sinal negativo. Ou seja, na história contada pelo modelo logístico, a escolaridade funciona apenas como *proxy* para a renda, esta sim protegendo seu detentor contra a violência mortal.

E agora?

Tendo em vista que histórias teóricas sustentando ambas as versões são possíveis, não nos resta outra alternativa que aprofundar a análise dos dados para ver se achamos outra pista. Os dados de renda no censo demográfico são coletados por entrevistadores treinados rapidamente com um questionário pouco detalhado. Mas a Pnad coleta esta variável com entrevistadores com maior treinamento e um questionário muito mais detalhado. Portanto, há pouca dúvida de que os dados de renda da Pnad são superiores aos do censo. A amostragem da Pnad também é suficiente para as 288 células permitidas pelo SIM. Sua única desvantagem se refere aos problemas de projeção populacional na Pnad, na qual a idade não entra. No entanto, acreditamos que o ajuste aqui feito, e descrito na subseção sobre metodologia (subseção 3.2), é adequado para resolver esta limitação, principalmente se o ano usado não for muito distante do ano do Censo de 2000. Usamos, pois, a Pnad de 2001, e os resultados se encontram dispostos no painel 2. A tabela completa se encontra anexa a este texto.

O painel 2 mostra que, usando uma medida melhor de renda, há alguma convergência entre as histórias contadas pelas duas metodologias, mas a aquela contada pelo modelo logístico continua, por assim dizer, “sem pé nem cabeça”. Embora a inclusão da renda ainda reduza o efeito da escolaridade no modelo logístico, os diferenciais entre 1 a 3 anos e as outras categorias ainda se mantêm fortes – mais ou menos o dobro do risco por ser negro, por exemplo. Contudo, o coeficiente da renda passa a ser positivo – ou seja, a renda passa a aumentar, significativo a 3%, as chances de o indivíduo morrer vítima de homicídio, embora esse coeficiente permaneça próximo de zero. A combinação de coeficiente de renda positivo e redução do coeficiente da educação, quando a renda é incluída, não é coerente com qualquer história teórica.

PAINEL 2

**Resultados de regressão – coeficientes de escolaridade em 2001**

GRÁFICO 1

**Modelo linear**

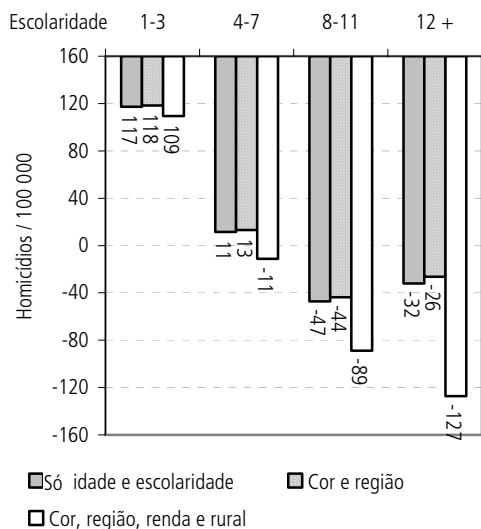
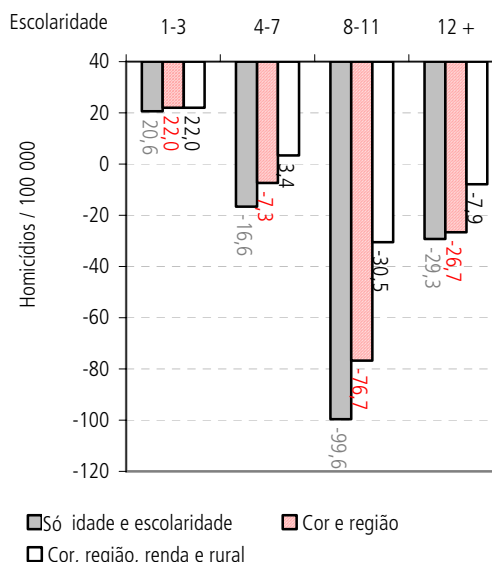


GRÁFICO 2

**Modelo logístico**



Fonte: SIM (2001) e Pnad (2001) – não inclui a região Norte.

Nota: Efeitos marginais na regressão logística calculados na taxa de homicídio: média de 2001.

O modelo linear conta exatamente a mesma história que antes. A renda aumenta o risco de homicídio, e a instrução o reduz fortemente. A maior diferença é que o efeito negativo de ter de 1 a 3 anos de estudo aumenta consideravelmente.

Por fim, apresentamos um painel com o comportamento dos cinco coeficientes de escolaridade para o modelo com renda e taxa de urbanização de 1999 a 2004. É difícil fazer uma inferência sem ambigüidade do comportamento dos coeficientes, uma vez que a qualidade dos dados tem mudado ao longo do tempo: a variável renda é medida com menor precisão em 2000, enquanto diversas variáveis do SIM, inclusive escolaridade, melhoram a partir de 2001. É importante lembrar que os coeficientes expressam o impacto de um dado nível de escolaridade com relação a nenhum estudo, ou seja, o coeficiente de zero ano de estudo é sempre o eixo horizontal.

O painel mostra uma estabilidade, a partir de 2001, do impacto da escolaridade sobre a probabilidade de morte por homicídio. A diferença na probabilidade de morte por homicídio entre indivíduos com 1 a 3 anos de estudo e aqueles com nível superior completo ou incompleto fica em média em 270 por 100 mil – uma diferença muito grande. Com relação aos que têm secundário completo ou incompleto, a diferença situa-se próxima a 217 por 100 mil – também muito grande.

Em qual modelo acreditar? Apesar de, em princípio, o modelo logístico ser mais adequado para investigar eventos raros como o homicídio, dado que há forte variação nas taxas de homicídio por célula, o mesmo talvez ofereça maior resolução para células cuja taxa é relativamente baixa, enquanto o modelo linear oferece maior resolução nas células cuja taxa é relativamente alta – que são as que mais interessam aqui. O modelo linear também é mais robusto com relação às hipóteses necessárias para que seja válido. Ademais, os resultados incoerentes do modelo logístico conferem mais

crédito ao modelo linear, cujos coeficientes contam uma história coerente entre si e também se mantêm mais ou menos coerentes ao longo do tempo.

PAINEL 3

### Resultados de regressão – coeficientes de escolaridade de 1999 a 2004

GRÁFICO 1

#### Modelo linear

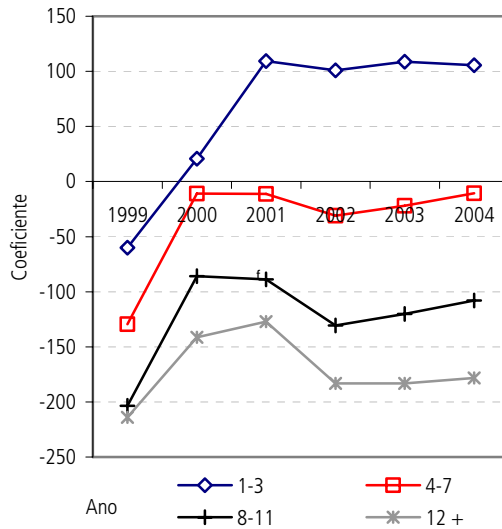
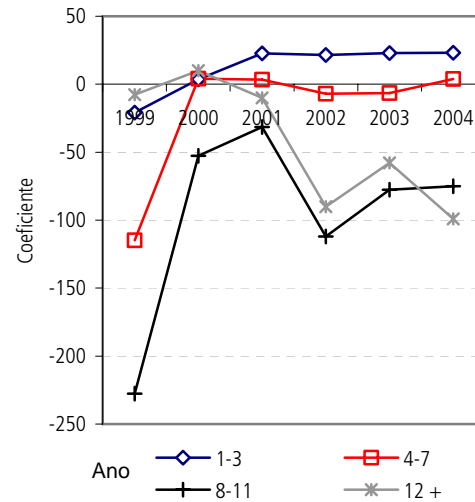


GRÁFICO 2

#### Modelo logístico



Fontes: SIM (2001) e Phad (2001).

Nota: Efeitos marginais na regressão logística calculados na taxa de homicídio: média de 1999-2004.

## 5 DISCUSSÃO E CONCLUSÃO

Se os coeficientes aqui estimados podem ser interpretados como relações causais, então os resultados deste texto são muito relevantes. A análise dos determinantes da taxa de homicídio torna visível que três variáveis exercem influência superior a 100 por cem mil: sexo, idade e nível de escolaridade. Um homem de trinta anos com renda *per capita* de R\$ 800 tem uma chance de 127 por 100 mil de morrer vítima de homicídio, contra nove para uma mulher com as mesmas características. A diferença nas taxas de mortalidade entre um homem de 50 anos ou mais e outro de 21 anos também é de 127 mortos por 100 mil. Por fim, a diferença na probabilidade de morrer vítima de homicídio entre um homem com 1 a 3 anos de estudo e outro com ensino superior é de 236 por 100 mil.

O impacto marginal de uma variável sobre a taxa de homicídio depende do produto da própria variável e seu coeficiente:  $\beta x$  no modelo linear e  $F(\beta x)$  no logístico. Isto sugere que há dois modos relacionados a uma dada variável para reduzir a taxa de homicídios. O primeiro é focar nos coeficientes, e há uma miríade de políticas de prevenção ao crime focadas em gênero, em jovens, ou nos que contam com escolaridade baixa. Ao se promover, por exemplo, um programa de esportes para jovens em uma comunidade carente, estão sendo utilizados esses resultados com vistas a mudar os coeficientes de idade e, possivelmente, de instrução.

O segundo modo é mudar o valor médio da variável. Dado que não é viável mudar o sexo das pessoas nem envelhecê-las, a única política pública passível de forte atuação sobre estes fortes determinantes do homicídio é a educação.

### ***Um escudo educacional?***

Mantendo ainda a hipótese de que os coeficientes representam relações causais, qual seria uma possível explicação para esta relação? Há décadas se sabe existir, na educação,<sup>15</sup> um forte conteúdo econômico que hoje se encontra concentrado no final do processo: o rendimento real de indivíduos de 30 a 40 anos com primário incompleto, primário completo, segundo grau completo e alguma instrução superior em 2001 era de R\$ 185, R\$ 260, R\$ 502 e R\$ 1.589, respectivamente. Ou seja, o conteúdo econômico é muito maior quando se completa o ensino médio e muito maior ainda quando se obtém alguma instrução superior. Trabalhos com gêmeos separados ao nascer e experimentos naturais como o serviço militar ou data de nascimento também deixam clara a direção da causalidade educação → renda. Sabe-se ainda que a relação entre este conteúdo econômico revelado pelo mercado de trabalho e o aprendizado de linguagem e matemática tem relação significativa, porém não muito forte.<sup>16</sup> Ou seja, a maior parte do conteúdo no processo educativo não é medido por exames padronizados de linguagem e matemática.

O processo educacional é composto por inúmeras interações que ocorrem tanto dentro como fora da escola, mas que por esta são mediadas. Sabemos pouco sobre muito do que ocorre neste processo e suas conseqüências no mundo exterior. Uma hipótese é que exista no processo educacional um conteúdo de convivência, ou até cidadania, que não tem valor econômico direto, mas que oferece proteção contra a violência. Como o processo educacional e o criminal são muito mal compreendidos pelos cientistas sociais, é difícil especular sobre qual seria a forma deste conteúdo. No entanto, imaginamos que talvez seja fruto do fato de interagir com outras crianças em um ambiente no qual, bem ou mal, o conflito é mediado. Desse modo as crianças aprenderiam a lidar com o conflito de forma a que o mesmo não leve à violência e, no fim da linha, à morte por homicídio. Se assim o for, é possível que a escola forneça um escudo educacional que proteja principalmente quem o detém, mas também terceiros que possam vir a entrar em situações de conflito com o detentor do escudo.

### ***A endogeneidade – será mesmo causal?***

O problema que assola econométristas desde que se descobriu que correlação não é causalidade é a endogeneidade. Define-se estatisticamente endogeneidade como sendo uma correlação entre uma variável explicativa e o resíduo. Mais intuitivo é definir endogeneidade como sendo um erro na definição do modelo que leva a um viés nos resultados. Ou seja, se alguma variável relevante foi esquecida, se há seleção na amostra, ou se há uma relação de causalidade reversa, os coeficientes estimados não representarão uma relação causal.

---

15. A literatura sobre este tema é extensa. Ver, por exemplo, Langoni (1973), Barros e Mendonça (1995) e Barros, Mendonça e Henriques (2001).

16. Ver Bowles, Gintis e Osborne (2001) para um resumo de literatura. Estes autores argumentam que o baixo poder explicativo do conhecimento de conteúdo como preditor do rendimento no mercado de trabalho se deve ao fato do aprendizado de atitudes, normas sociais e valores ser mais importante que o domínio de conteúdo.

No caso das regressões aqui descritas é possível que haja uma variável omitida – por exemplo, fragilidade familiar. Esta variável pode levar tanto a menor desempenho educacional quanto a maior propensão a morrer vítima de homicídio, gerando assim uma correlação espúria entre os dois. Suponhamos, à guisa de exemplo, que uma família cai na pobreza e que se verifique também desintegração do núcleo familiar. Isto pode levar as crianças a saírem da escola por falta de apoio familiar ao estudo, assim como também pode levar algumas crianças a entrarem em atividades criminosas que as expõem ao risco de homicídio, o que implicaria uma correlação não causal entre instrução e risco de homicídio em função da omissão de uma variável. Não é difícil acreditar que este seja o caso.

É possível ainda contar outras histórias nas quais a relação entre instrução e taxa de homicídio seja mais complexa que a equação cujos coeficientes se encontram na tabela 2. Uma possibilidade é que a taxa de subdeclaração de níveis de instrução baixos seja maior que a taxa de níveis de instrução altos, mesmo condicionados à idade, município e data de ocorrência. Se assim o for, o procedimento *hot deck* usado subestimarà o número de homicídios de indivíduos pouco instruídos e o sobreestimarà entre os muito instruídos. Em todos estes casos, os coeficientes serão endógenos por viés de seleção e não podem ser interpretados como causais.

A solução, como sempre, se encontra na identificação do impacto dessa variável mediante sua inclusão explícita na equação, ou a exclusão do seu efeito mediante o uso de uma variável instrumental para nível de instrução. Para a primeira estratégia seria necessário ter uma variável que servisse como *proxy* de origem socioeconômica, o que não é possível com os dados dos quais dispomos. Para a segunda, será necessário encontrar uma variável que seja determinante do nível educacional, mas não da taxa de homicídio, que permita filtrar efeitos da variável omitida. Acreditamos que tais variáveis existam, mas trata-se de uma tarefa complexa a ser abordada em outro trabalho.

### ***Implicações para a política de segurança pública e educacional***

Ainda que se possa fazer a interpretação causal dos coeficientes aqui estimados, este texto não contém muitas implicações para a política de segurança pública. Assuntos como o tamanho ótimo das penas, redução da maioria penal, o modelo de policiamento, integração das polícias, fluxo criminal e política carcerária são ortogonais ao impacto da educação sobre a propensão a ser vítima de homicídio. O quanto penas mais severas levam à dissuasão de atividade criminosa ou se o modelo de policiamento por ocorrência é adequado às circunstâncias brasileiras podem até ser influenciados pelo nível educacional da população, mas este texto não investiga esta relação. A única conclusão – tentativa – é que a educação formal parece ter um efeito redutor muito forte sobre a taxa de homicídio, e que isto possivelmente se deva ao papel socializador da escola.

Entretanto, as conclusões para a política educacional são extremamente fortes. O texto oferece uma evidência substantiva a favor de manter as crianças na escola, mesmo se a aprendizagem de conteúdos ficar abaixo das expectativas, já reduzidas, da sociedade. Há um discurso recorrente contra políticas educacionais que visam à permanência, tais como ciclos educacionais, e até a sua versão mais radical: a progressão continuada. Frequentemente, este discurso se vale do argumento de que nada adianta aprovar crianças que não estão aprendendo ou que é inútil formar

analfabetos. Acreditamos que este argumento falha grosseiramente na medida em que não especifica o que não está sendo aprendido. Este texto mostra evidências de que, mesmo que uma criança de baixo status socioeconômico freqüentando uma escola com professores mal pagos e mal formados não esteja aprendendo português ou matemática a contento, ela está aprendendo um modo de socialização que eventualmente poderá salvar-lhe a vida. E mais: é possível que, ao ensinar esta criança a como lidar com o conflito de modo não letal, a escola esteja também salvando a vida de terceiros.

A conclusão inexorável é que a política educacional deve fazer tudo ao seu alcance para manter a criança na escola, mesmo que a aprendizagem de conteúdos acadêmicos seja aquém do desejado. Nesse sentido, políticas de progressão continuada devem ser incentivadas ao máximo, uma vez que há uma relação conhecida entre ser reprovado e evadir do processo educacional.<sup>17</sup> Não se trata apenas de aprender a ler e escrever: é questão de vida e morte.

---

17. Ver Ribeiro (1991).

## REFERÊNCIAS

- BARROS, M. D. de A.; XIMENES, R.; LIMA, M. L. C. de. Mortalidade por causas externas em crianças e adolescentes: tendências de 1979 a 1995. *Revista de Saúde Pública*, São Paulo, v. 35, n. 2, 2001. Disponível em: <<http://www.scielo.br/scielo>>.
- BARROS, R. P.; MENDONÇA, R. *Os determinantes da desigualdade no Brasil*. Rio de Janeiro, jul. 1995 (Texto para Discussão, n. 377). Disponível em: <<http://www.ipea.gov.br/>>.
- BARROS, R. P.; MENDONÇA, R.; HENRIQUES, R. *A estabilidade inaceitável: desigualdade e pobreza no Brasil*. Rio de Janeiro, jun. 2001 (Texto para Discussão, n. 800). Disponível em: <<http://www.ipea.gov.br/>>.
- BATISTA, L. E. Masculinity, race/color and health. *Ciência & saúde coletiva*, v. 10, n. 1, p. 71-80, 2005. Disponível em: <<http://www.scielo.br/scielo>>.
- BOWLES, S.; GINTIS, H.; OSBORNE, M. The determinants of earnings: a behavioral approach. *Journal of Economic Literature*, v. 39, n. 4, p. 1.137-1.176, Dec. 2001.
- CARVALHO, A.; CERQUEIRA, D.; LOBÃO, W. *Socioeconomic structure, self-fulfillment, homicides and spatial dependence in Brazil*. Rio de Janeiro, jul. 2005 (Texto para Discussão, n. 1.105). Disponível em: <<http://www.ipea.gov.br/>>.
- FERREIRA, H.; ARAUJO, H. E. Transições negadas: homicídios entre os jovens brasileiros. In: CAMARANO, A. A. (Org.). *Transição para a vida adulta ou vida adulta em transição?* Rio de Janeiro: Ipea, 2006.
- GROGGER, J. Market wages and youth crime. *Journal of Labor Economics*, v. 16, n. 4, p. 756-791, Oct. 1998.
- GUTIERREZ, M. B. S. *et al. Inequality and criminality revisited: further evidence from Brazil*. Disponível em: <[http://www.anpec.org.br/encontro\\_2004.htm](http://www.anpec.org.br/encontro_2004.htm)>.
- HARLOW, C. W. *Education and correctional populations*. Bureau of Justice Statistics Special Report, 2003. Disponível em: <<http://www.ojp.usdoj.gov/bjs/abstract/ecp.htm>>.
- HU, M.; SALVUCCI, S. M.; COHEN, M. P. *Evaluation of some popular imputation algorithms*. Proceedings of the Section on Survey Research Methods. American Statistical Association. 1998. Disponível em: <[http://www.amstat.org/Sections/Srms/Proceedings/papers/1998\\_049.pdf](http://www.amstat.org/Sections/Srms/Proceedings/papers/1998_049.pdf)>.
- IMAI, S.; KATAYAMA, H.; KRISHNA, K. *Crime and young men: the role of arrest, criminal experience, and heterogeneity*. May 2006 (NBER Working Paper, n. 12.221). Disponível em: <<http://www.nber.org/>>.
- JOHNSON, R. S.; KANTOR, S.; FISHBACK, P. V. *Striking at the roots of crime: the impact of social welfare spending on crime during the Great Depression*. 2007 (NBER Working Paper, n. 12.825). Disponível em: <<http://www.nber.org/>>.
- KILSZTAJN, S. *et al. Vítimas da cor: homicídios na região metropolitana de São Paulo, Brasil, 2000*. *Cadernos de Saúde Pública*, v. 21, n. 5, p. 1.408-1.415, out. 2005. Disponível em: <<http://www.scielo.br/scielo>>.
- LANGONI, C. G. *Distribuição de renda e desenvolvimento econômico do Brasil*. Rio de Janeiro: Expressão e Cultura, 1973.
- LOCHNER, L.; MORETTI, E. *The effect of education on crime: evidence from prison inmates, arrests, and self-reports*. 2001 (NBER Working Paper, n. 8.605). Disponível em: <<http://www.nber.org/>>.



- MACEDO, A. C. *et al.* Violência e desigualdade social: mortalidade por homicídios e condições de vida em Salvador, Brasil. *Revista de Saúde Pública*, v. 35, n. 6, p. 515-522, dez. 2001. Disponível em: <<http://www.scielo.br/scielo>>.
- MADDALA, G. S. Limited-dependent and qualitative variables in econometrics. *Econometric Society Monographs*, n. 3. Cambridge University Press, 1983.
- MOCAN, H. N.; REES, D. I. *Economic conditions, deterrence and juvenile crime: evidence from micro data*. Oct. 1999 (NBER Working Paper, n. 7.405). Disponível em: <<http://www.nber.org/>>.
- NÉRI, M. *Retratos do cárcere*. Mimeo Centro de Políticas Sociais/FGV, 2006.
- OLIVEIRA, C. A. *Criminalidade e o tamanho das cidades brasileiras: um enfoque da economia do crime*. Anpec, 2005. Disponível em: <[http://www.anpec.org.br/encontro\\_2005.htm](http://www.anpec.org.br/encontro_2005.htm)>.
- RIBEIRO, S. C. A pedagogia da repetência. *Estudos em Avaliação Educacional*, São Paulo, n. 4, p. 73-86, jul./dez. 1991. Disponível em: <<http://www.scielo.br>>.
- SEASTROM, M.; KAUFMAN, S.; LEE, R. *Evaluating the impact of imputations for item nonresponse*. NCES Statistical Standards. Ed por U.S. Department of Education. Institute of Education Sciences Washington, D. C., 2003. Disponível em: <<http://nces.ed.gov>>.
- SOARES FILHO, A. M. *et al.* Análise da mortalidade por homicídio no Brasil. Epidemiologia e serviços de saúde: *Revista do sistema único de saúde no Brasil*, v. 16, n. 1. jan./mar. 2007. Disponível em: <<http://bvsmms.saude.gov.br/html/pt/periodicos.html>>.
- UNICEF. FUNDO DAS NAÇÕES UNIDAS PARA A INFÂNCIA. *Retrato estatístico das mortes de crianças e jovens por causas violentas: Brasil 1979-1993*. Brasília, 1995.
- WAISELFISZ, J. J. *Mapa da violência II: os jovens do Brasil*. Rio de Janeiro: Unesco, Instituto Ayrton Senna, Ministério da Justiça, 2000.
- \_\_\_\_\_. *Mapa da violência III*. Brasília :Unesco, Instituto Ayrton Senna, Ministério da Justiça/Secretaria Especial dos Direitos Humanos, 2002.
- \_\_\_\_\_. *Mapa da violência VI*. Brasília :Unesco, Instituto Ayrton Senna, Ministério da Justiça/Secretaria Especial dos Direitos Humanos, 2004.
- \_\_\_\_\_. *Mapa da violência: os jovens do Brasil*. Rio de Janeiro: Unesco, Instituto Ayrton Senna, Garamond, 1998.
- WITTE, A. D.; TAUCHEN, H. *Work and crime: an exploration using panel data*. Jul. 1994 (NBER Working Paper, n. 4.194). Disponível em: <<http://www.nber.org/>>.

# ANEXO

TABELA A.1

## Resultados de regressão para 2001 – Modelo Linear de Probabilidade

		Modelo Linear de Probabilidade					
		$\beta$	p-valor	$\beta$	p-valor	$\beta$	p-valor
<b>Homens</b>							
Idade							
	Idade 10	0.00		0.00		0.00	
	Idade 15	55.13	1%	53.17	1%	60.41	0%
	Idade 18	110.85	0%	109.71	0%	117.72	0%
	Idade 21	122.34	0%	121.71	0%	134.01	0%
	Idade 24	122.96	0%	121.05	0%	126.17	0%
	Idade 27	116.68	0%	114.72	0%	112.67	0%
	Idade 30	107.63	0%	105.31	0%	96.29	0%
	Idade 33	93.77	0%	90.81	0%	76.57	0%
	Idade 36	86.25	0%	83.90	0%	70.49	0%
	Idade 40	72.68	0%	69.78	0%	47.35	2%
	Idade 50	41.56	2%	38.31	3%	5.99	76%
Escolaridade							
	zero	0.00		0.00		0.00	
	1-3	117.13	0%	118.21	0%	109.44	0%
	4-7	11.43	41%	13.10	33%	-11.14	59%
	8-11	-47.08	0%	-43.73	0%	-88.92	0%
	12 +	-32.14	5%	-26.42	9%	-127.35	0%
Cor/raça							
	Negra			26.11	0%	31.55	0%
	Branca			0.00		0.00	
Região							
	Nordeste			-27.29	0%	-17.70	9%
	Sudeste, Sul, Centro-Oeste			0.00		0.00	
Renda domiciliar pc (R\$ 100)							
						4.19	0%
Rural							
						-65.68	40%
Constante		-18.99	34%	-17.39	38%	9.93	76%
<b>Mulheres</b>							
Idade							
	Idade 10	0.00		0.00		0.00	
	Idade 20	4.77	3%	4.77	3%	5.42	1%
	Idade 30	4.46	4%	4.41	4%	4.24	5%
	Idade 40	3.15	15%	2.89	17%	2.01	36%
	Idade 50	0.31	88%	0.18	93%	-1.42	55%
Escolaridade							
	zero	0.00		0.00		0.00	
	1-3	7.61	0%	7.91	0%	7.69	0%
	4-7	-0.05	97%	0.34	81%	-0.19	93%
	8-11	-3.36	2%	-3.10	3%	-4.65	19%
	12 +	-1.69	30%	-1.15	48%	-6.74	24%
Cor/raça							
	Negra			1.32	14%	1.67	7%
	Branca			0.00		0.00	
Região							
	Nordeste			-2.05	2%	-1.83	20%
	Sudeste, Sul, Centro-Oeste			0.00		0.00	
Renda domiciliar pc (R\$ 100)							
						0.56	10%
Rural							
						0.19	99%
Constante		1.73	45%	1.90	41%	2.00	60%

Fontes: SIM (2001) e Phad (2001) – exclui a região Norte.

TABELA A.2

**Resultados de regressão para 2001 – Modelo Logístico**

		Regressão Logística					
		$\beta$	p-valor	$\beta$	p-valor	$\beta$	p-valor
<b>Homens</b>							
Idade							
	Idade 10	0.0		0.0		0.0	
	Idade 15	23.7	0%	23.6	0%	23.5	0%
	Idade 18	28.4	0%	28.5	0%	28.4	0%
	Idade 21	27.8	0%	28.0	0%	28.1	0%
	Idade 24	28.1	0%	28.3	0%	28.2	0%
	Idade 27	28.3	0%	28.4	0%	28.3	0%
	Idade 30	28.2	0%	28.3	0%	28.1	0%
	Idade 33	27.6	0%	27.7	0%	27.4	0%
	Idade 36	27.4	0%	27.2	0%	27.0	0%
	Idade 40	26.3	0%	26.4	0%	25.8	0%
	Idade 50	24.0	0%	23.8	0%	22.5	0%
Escolaridade							
	zero	0.0		0.0		0.0	
	1-3	20.6	0%	22.0	0%	22.0	0%
	4-7	-16.6	2%	-7.3	11%	3.4	48%
	8-11	-99.6	0%	-76.7	0%	-30.5	0%
	12 +	-29.3	0%	-26.7	0%	-7.9	44%
Cor/raça							
	Negra			16.7	0%	17.0	0%
	Branca			0.0		0.0	
Região							
	Nordeste			-20.4	0%	-32.9	0%
	Sudeste, Sul, Centro-Oeste			0.0		0.0	
Renda domiciliar pc (R\$ 100)							
						28.9	3%
Rural							
						26.8	0%
Constante		7.1	0%	6.4	0%	3.5	0%
<b>Mulheres</b>							
Idade							
	Idade 10	0.0		0.0			
	Idade 20	18.2	1%	18.8	0%	19.0	0%
	Idade 30	16.3	2%	17.0	0%	17.2	0%
	Idade 40	10.9	16%	11.4	7%	12.6	5%
	Idade 50	-5.0	61%	-4.0	62%	0.2	98%
Escolaridade							
	zero	0.0		0.0		0.0	
	1-3	18.9	0%	20.8	0%	20.6	0%
	4-7	-30.1	2%	-18.5	5%	-4.9	64%
	8-11	-52.2	0%	-45.8	0%	-13.0	44%
	12 +	-20.5	5%	-14.1	8%	7.0	70%
Cor/raça							
	Negra			13.0	0%	12.9	0%
	Branca			0.0		0.0	
Região							
	Nordeste			-27.2	0%	-45.4	0%
	Sudeste, Sul, Centro-Oeste			0.0		0.0	
Renda domiciliar pc (R\$ 100)							
						0.2	86%
Rural							
						27.7	12%
Constante		2.7	0%	2.8	0%	1.5	0%

Fontes: SIM (2001) e Pnad (2001) – exclui a região Norte.

TABELA A.3

**Coefficientes estimados: 1999 a 2004 – Modelo linear**

	1999	2000	2001	2002	2003	2004
<b>Homens</b>						
Idade						
Idade 10	0.00	0.00	0.00	0.00	0.00	0.00
Idade 15	41.70	69.31	61.63	71.59	71.48	70.59
Idade 18	82.37	114.00	118.78	120.84	130.44	123.76
Idade 21	97.82	134.95	135.33	141.24	159.06	154.64
Idade 24	96.97	129.41	126.86	136.04	144.70	140.12
Idade 27	76.78	118.08	113.52	108.51	120.74	117.23
Idade 30	58.80	110.50	97.28	95.66	97.54	89.27
Idade 33	47.93	103.34	77.52	69.92	74.86	68.58
Idade 36	39.85	101.23	71.50	65.57	61.80	57.61
Idade 40	22.19	82.91	48.40	45.58	43.23	39.69
Idade 50	-27.30	42.31	7.15	-3.28	-7.47	-6.02
Escolaridade						
zero	0.00	0.00	0.00	0.00	0.00	0.00
1-3	-60.00	20.61	109.47	100.88	108.63	105.50
4-7	-129.42	-10.88	-11.29	-31.05	-22.00	-10.62
8-11	-203.40	-85.94	-88.80	-130.55	-120.10	-108.00
12 +	-213.88	-141.37	-127.05	-183.22	-183.31	-178.21
Cor/raça						
Negra	24.26	29.09	31.96	35.01	39.63	38.97
Branca	0.00	0.00	0.00	0.00	0.00	0.00
Região						
Nordeste	-13.94	-12.41	-18.12	-6.50	-12.24	-7.46
Sudeste, Sul, Centro-Oeste	0.00	0.00	0.00	0.00	0.00	0.00
Renda domiciliar pc (R\$ 100)						
	4.00	8.05	4.20	4.40	4.96	4.70
Rural	-189.43	-28.21	-65.09	-194.64	-141.60	-152.38
Constante	165.92	-24.79	8.63	53.62	34.36	26.60
<b>Mulheres</b>						
Idade						
Idade 10	0.00	0.00	0.00	0.00	0.00	0.00
Idade 20	3.57	8.82	5.45	5.56	5.50	5.96
Idade 30	1.89	8.42	4.25	3.71	3.50	3.83
Idade 40	0.09	7.28	2.04	1.36	1.01	1.56
Idade 50	-4.43	3.20	-1.48	-2.68	-2.92	-2.40
Escolaridade						
zero	0.00	0.00	0.00	0.00	0.00	0.00
1-3	-3.76	1.84	7.61	6.62	7.10	7.49
4-7	-8.10	-0.35	-0.29	-1.88	-0.17	0.31
8-11	-13.17	-5.79	-4.83	-7.42	-5.77	-5.24
12 +	-12.99	-8.00	-6.96	-10.69	-8.62	-8.78
Cor/raça						
Negra	1.52	1.46	1.65	1.59	1.79	1.69
Branca	0.00	0.00	0.00	0.00	0.00	0.00
Região						
Nordeste	-1.54	-0.98	-1.76	-1.35	-1.84	-0.74
Sudeste, Sul, Centro-Oeste	0.00	0.00	0.00	0.00	0.00	0.00
Renda domiciliar pc (R\$ 100)						
	0.60	0.58	0.57	0.57	0.59	0.56
Rural	-10.35	0.63	-0.28	-6.28	-2.05	-3.97
Constante	13.21	-2.07	2.15	5.56	3.74	2.92

Fontes: SIM (1999-2004), Censo Demográfico (2000) e Pnads (1999 e 2001-2004) – todas excluem a região Norte.

TABELA A.4

**Coefficientes estimados: 1999 a 2004 – Modelo logístico**

	1999	2000	2001	2002	2003	2004
<b>Homens</b>						
Idade						
Idade 10	0.00	0.00	0.00	0.00	0.00	0.00
Idade 15	25.1	30.5	23.5	22.7	26.8	23.1
Idade 18	28.5	31.6	29.4	29.1	29.5	29.3
Idade 21	28.3	31.7	29.2	29.6	30.1	29.9
Idade 24	29.6	31.7	29.1	29.4	30.0	30.1
Idade 27	29.0	31.3	29.2	29.3	29.5	29.4
Idade 30	28.4	31.5	29.0	29.1	28.7	28.6
Idade 33	27.9	31.8	28.2	28.1	27.6	27.8
Idade 36	28.7	31.7	27.9	27.9	27.6	26.8
Idade 40	27.7	31.5	26.3	26.1	25.9	26.3
Idade 50	26.0	31.2	22.2	19.2	20.6	20.3
Escolaridade						
zero	0.00	0.00	0.00	0.00	0.00	0.00
1-3	-20.9	3.7	22.6	21.6	23.0	23.3
4-7	-114.9	4.0	3.3	-7.0	-6.5	3.7
8-11	-227.5	-52.8	-31.4	-111.9	-77.5	-75.0
12 +	-7.8	10.0	-10.0	-90.2	-57.8	-99.0
Cor/raça						
Negra	16.2	19.7	17.5	20.3	18.1	20.1
Branca						
Região						
Nordeste	-50.1	-57.0	-33.2	-28.3	-24.5	-32.5
Sudeste, Sul, Centro-Oeste						
Renda domiciliar pc (R\$ 100)	-1,114	-10,961	30.6	31.7	30.0	31.2
Rural	15.0	27.7	27.9	-2.7	-3.6	4.6
Constante	21.1	1.1	4.0	7.8	7.8	6.3
<b>Mulheres</b>						
Idade						
Idade 10	0.00	0.00	0.00	0.00	0.00	0.00
Idade 20	11.6	29.8	19.8	16.7	19.7	20.6
Idade 30	0.0	29.9	17.5	8.2	18.3	17.0
Idade 40	4.9	29.5	14.0	-2.8	4.2	11.5
Idade 50	-12.6	29.0	-4.0	-43.0	-0.1	-20.8
Escolaridade						
zero	0.00	0.00	0.00	0.00	0.00	0.00
1-3	-24.5	2.4	21.5	17.3	21.9	23.2
4-7	-77.5	8.8	-9.1	-32.0	6.2	-0.1
8-11	-101.7	-13.8	-23.3	-106.1	-49.6	-37.6
12 +	13.5	19.0	0.7	-71.8	-11.4	-24.1
Cor/raça						
Negra	12.1	11.5	13.0	13.7	10.9	16.6
branca						
Região						
Nordeste	-103.2	-43.7	-39.7	-34.0	-45.8	-23.9
Sudeste, Sul, Centro-Oeste						
Renda domiciliar pc (R\$ 100)	-0.02	-0.03	0.03	0.06	0.04	0.04
Rural	30.6	31.1	26.5	11.8	24.1	24.9
Constante	6.1	0.2	1.7	5.4	2.1	1.6

Fontes: SIM (1999-2004), Censo Demográfico (2000) e Phads (1999 e 2001-2004) – todas excluem a região Norte.



## **EDITORIAL**

### **Coordenação**

Iranilde Rego

### **Supervisão**

Aeromilson Mesquita

### **Revisão**

Marco Aurélio Dias Pires

Ângela Pereira da Silva Oliveira (estagiária)

Camila de Paula Santos (estagiária)

Nathalia Martins Peres da Costa (estagiária)

### **Editoração**

Bernar José Vieira

Elidiane Bezerra Borges

Iranilde Rego

### **Brasília**

SBS – Quadra 1 – Bloco J – Ed. BNDES, 9º andar

70076-900 – Brasília – DF

Fone: (61) 3315-5090

Fax: (61) 3315-5314

Correio eletrônico: editbsb@ipea.gov.br

### **Rio de Janeiro**

Av. Nilo Peçanha, 50, 6º andar – Grupo 609

20044-900 – Rio de Janeiro – RJ

Fone: (21) 3515-8433

Fax: (21) 3515-8402

Correio eletrônico: editrj@ipea.gov.br

## **COMITÊ EDITORIAL**

### **Secretário-Executivo**

Marco Aurélio Dias Pires

SBS – Quadra 1 – Bloco J – Ed. BNDES,  
9º andar, sala 908

70076-900 – Brasília – DF

Fone: (61) 3315-5406

Correio eletrônico: madp@ipea.gov.br

Tiragem: 130 exemplares