

Análise da situação ocupacional de crianças e adolescentes nas regiões Sudeste e Nordeste do Brasil utilizando informações da PNAD 1999

Phillippe George Pereira Guimarães Leite*
Denise Britz do Nascimento Silva**

Utilizando os dados coletados pela PNAD 1999, o artigo investiga a influência das características individuais e das famílias de crianças e adolescentes entre 5 e 17 anos residentes nas regiões Sudeste e Nordeste do Brasil sobre as escolhas ocupacionais: somente participar da força de trabalho, somente estudar, realizar ambas as atividades ou não executar nenhuma atividade. Todas as estimações necessárias para o estudo foram calculadas incorporando o plano amostral da PNAD. Como conclusão, observou-se que a atividade da criança está relacionada com as condições de vida e o nível de pobreza aos quais está submetida. Quanto menos instruída for sua mãe, maior a chance de uma criança abandonar a escola, sendo que viver no Nordeste implica uma maior chance de realizar ambas as atividades, ao passo que residir no Sudeste implica uma maior chance de permanecer fora da força de trabalho.

Introdução

Durante a última década, o número de ações e de programas que buscam a erradicação do trabalho prejudicial ao desenvolvimento das crianças e dos adolescentes aumentou consideravelmente no Brasil. Ainda assim, estatísticas oficiais fornecem evidências sobre a dificuldade em alcançarmos uma solução concreta e definitiva para tal problema. Estima-se, com base nos dados coletados pela Pesquisa Nacional por Amostra de Domicílios (PNAD) de 1999, que 3,2 milhões de indivíduos com idade entre 5 e 14 anos participavam da força de trabalho no Brasil naquele ano, representando 2% da população¹ total estimada.

Os estudiosos apontam a complementação da renda domiciliar como sendo o principal motivo da inserção da criança no mercado de trabalho. As camadas economicamente menos favorecidas da

sociedade constituem um grupo social caracterizado por um nível educacional muito baixo e por um alto índice de desemprego. O baixo nível educacional, devido, principalmente, ao abandono escolar, pode ser associado à falta de instituições de ensino público ou mesmo à baixa qualidade do ensino prestado.

A ausência de professores qualificados ou motivados, o déficit de vagas nas instituições próximas à residência e a falta de material escolar (por exemplo, mesas e cadeiras para os alunos, bem como livros didáticos) são problemas notórios em escolas localizadas nas áreas mais carentes e fora dos grandes centros urbanos do país. Nas comunidades carentes dos grandes centros urbanos é possível observar, também, a ocorrência do abandono escolar devido ao recrutamento de crianças e adolescentes para a execução

* Mestre em Estudos Populacionais e Pesquisas Sociais pela Escola Nacional de Ciências Estatísticas (ENCE/IBGE), pesquisador do Departamento de Economia da Pontifícia Universidade Católica do Rio de Janeiro, da ENCE/IBGE e do Banco Mundial.

** Ph.D. em Estatística pela Universidade de Southampton e professora da Escola Nacional de Ciências Estatísticas (ENCE/IBGE).

¹ Estimada em 160 milhões de pessoas, de acordo com os dados da PNAD/IBGE 1999.

de atividades relacionadas ao tráfico de drogas (cf. Misse, 1997). Entretanto, a permanência de uma criança ou adolescente na escola é determinada sobretudo pela decisão dos pais. Este fato é observado inclusive nas camadas mais pobres da sociedade, mesmo quando as instituições de ensino não são atrativas (cf. Souto de Oliveira, 1994).

Nem sempre as crianças e os adolescentes abandonam a escola para participar do mercado de trabalho. Uma parcela significativa destes está fora da escola e fora da força de trabalho. No ano de 1999, pertenciam a este grupo 2,6 milhões de crianças e adolescentes entre 5 e 14 anos, representando 1,6% da população total do Brasil (PNAD/IBGE, 1999). Comparando este quantitativo ao total de indivíduos da mesma faixa etária que participavam da força de trabalho (3,2 milhões), nota-se que não só o trabalho infantil deve ser tratado como fonte de preocupação no que diz respeito à questão da criança e do adolescente, mas também o excedente de crianças e adolescentes que não são inseridas nem no universo escolar, nem na força de trabalho.

Considerando que a pobreza e a concentração de riqueza são alguns dos principais fatores que contribuem para a existência do trabalho infantil, este artigo concentra a análise dos dados em apenas duas regiões geográficas brasileiras, o Sudeste e o Nordeste, regiões altamente desiguais em termos de renda e nível de desenvolvimento. A base de dados utilizada é proveniente da Pesquisa Nacional por Amostra de Domicílios (PNAD) realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) no ano de 1999. O grupo etário de interesse é composto por todas as crianças e adolescentes de 5 a 17 anos residentes nas regiões geográficas definidas acima. Para caracterizar a situação ocupacional foram consideradas quatro categorias ocupacionais mutuamente exclusivas, a saber:

- a) somente participa da força de trabalho;
- b) somente estuda;
- c) participa da força de trabalho e estuda;
- d) não participa da força de trabalho nem estuda.

Com o intuito de analisar simultaneamente as quatro categorias, aplica-se um modelo multinomial logístico condicionado nas características individuais da criança, características dos pais e características do domicílio em que residem. Adicionalmente, a modelagem estatística desenvolvida neste artigo incorpora o desenho amostral da PNAD em seu procedimento de estimação, reconhecendo que a pesquisa é definida sob um plano amostral complexo. Cabe ressaltar que este tema já foi estudado por Leme e Wajnman (2000) e por Corseuil, Foguel e Santos (2001), mas sem a incorporação do plano amostral da pesquisa utilizada.

Apresenta-se, a seguir, o conceito de trabalho infantil e as leis nacionais que o definem. Na terceira seção são descritas as características e o plano amostral da PNAD/IBGE e na quarta seção são definidos os conceitos para a análise de dados amostrais complexos. A teoria referente ao modelo multinomial logístico empregado na análise dos dados da PNAD é descrita na seção que se segue. Finalmente, na sexta seção encontram-se as análises e, na sétima, a conclusão do artigo.

Trabalho infantil

O trabalho infantil é associado à pobreza e esta, por sua vez, a um processo histórico de crescimento econômico, não sendo decorrente das características individuais ou morais dos indivíduos. Nos países em desenvolvimento, e nos subdesenvolvidos, observam-se um maior número de pessoas vivendo em condições de pobreza e altas taxas de participação das crianças e adolescentes na força de trabalho. O mercado de trabalho disponível para os mais pobres é, em geral, restrito a ocupações cuja remuneração mensal é muito baixa. Há casos em que o exercício desta atividade ocorre em locais não propícios, estando os trabalhadores sujeitos a acidentes de trabalho e, em muitos casos, colocando suas vidas, ou saúde, em risco permanente. Além disso, alguns trabalhos podem ter características de exploração, tais como: *muitas horas de atividade*;

atividade e vida nas ruas; remuneração inadequada; atividade que impede o acesso à educação; atividade que compromete a dignidade e a auto-estima, como escravidão, trabalho servil e exploração sexual; atividade prejudicial ao pleno desenvolvimento social e psicológico.

Trabalhos com tais características, quando realizados por crianças ou adolescentes, determinam as circunstâncias do trabalho infantil (trabalho penoso, prejudicial ao desenvolvimento da criança e do adolescente). Nas zonas urbanas, crianças e adolescentes perambulam pelas ruas vendendo balas, engraxando sapatos ou empurrando carrinhos de feira, enquanto nas áreas rurais trabalhadores infantis freqüentemente não são remunerados monetariamente e, em muitos casos, trabalham ou ajudam na colheita para a própria subsistência.

Por outro lado, o trabalho realizado por crianças e adolescentes pode não ser penoso e não ter características de exploração, colaborando para o desenvolvimento de novas habilidades e do senso de responsabilidade das crianças e adolescentes. Para ser aceito pela sociedade, é necessário que este trabalho vise ao bem-estar das crianças e ao acesso à educação². A Constituição Federal Brasileira proíbe o trabalho a menores de 16 anos, exceto para aqueles que trabalham na condição de aprendizes, cuja idade mínima é 14 anos.

Nos últimos anos, observou-se a criação de inúmeros projetos destinados à erradicação do trabalho infantil no Brasil. Estes projetos são voltados para a distribuição de recursos às famílias que possuem crianças e adolescentes trabalhando, pois consideram que a insuficiência de renda é um dos principais fatores de inserção da criança e do adolescente no mercado de trabalho. A condição de dependência das famílias com relação às suas crianças para a geração de renda é o ponto inicial do

chamado ciclo de pobreza, no qual as famílias pobres consideram o trabalho infantil como sendo, talvez, a única válvula de escape contra a fome e a miséria.

Em alguns casos, a remuneração obtida pela criança ou pelo adolescente representa a única fonte de renda das famílias ou constitui parte significativa do orçamento familiar. A inserção no mercado de trabalho ocorre por influência direta dos pais, na busca de complementação do orçamento doméstico, transformando o trabalho das crianças numa solução de curto prazo para os problemas das famílias.

As crianças e os adolescentes que deixam de estudar para participar do mercado de trabalho estão, no futuro, sujeitos a ocupações de baixa remuneração³ e a condições precárias de trabalho. A falta de oportunidade, assim como de perspectiva no futuro, induz a formação de um ciclo penoso para a criança que se insere no mercado de trabalho precocemente. No futuro, este "adulto", ao constituir sua família, continuará vivendo sob as mesmas condições de penúria observadas quando criança, estabelecendo-se um *ciclo de pobreza*, reflexo da falta de oportunidades e de sua baixa qualificação profissional.

[...] a baixa escolaridade, causada pela entrada precoce no mercado de trabalho, tem o efeito de limitar as oportunidades de emprego a postos de trabalho que não exigem qualificação e que dão baixa remuneração, mantendo o jovem dentro de um ciclo de pobreza já experimentado pelos pais. (Silva e Kassouf, 2001)

A Pesquisa Nacional por Amostra de Domicílios – PNAD/IBGE

A Pesquisa Nacional por Amostra de Domicílios (PNAD/IBGE) tem como objetivo produzir informações que auxiliem o estudo do desenvolvimento socioeconômico do Brasil. O plano amostral da pesquisa

² Por exemplo, projetos realizados por empresas, ou desenvolvidos por agremiações culturais e esportivas, nos quais a participação em atividades esportivas ou profissionalizantes é condicionada a um bom aproveitamento escolar.

³ Estudos relacionam uma melhor remuneração mensal a um maior nível educacional e de qualificação profissional (Barros, Henriques e Mendonça, 2000). Os trabalhadores não qualificados ingressam no mercado de trabalho da informalidade, principalmente no setor de serviços, estando sujeitos às dificuldades inerentes à precarização do mercado de trabalho.

permite a obtenção de resultados para a totalidade do país⁴, para cada uma das regiões geográficas e para cada estado brasileiro. A amostra tem um desenho complexo em múltiplos estágios denominado Amostragem Estratificada de Conglomerados em Múltiplos Estágios de Seleção.

A primeira etapa para a composição do desenho amostral é a definição dos estratos. Para isso, os municípios são classificados como auto-representativos (AR)⁵ e não auto-representativos (NAR). Cada município AR é considerado um estrato, ao passo que os municípios NAR são agrupados em microrregiões que constituem estratos. Dentro de cada estrato são selecionadas as unidades primárias de amostragem (municípios), com probabilidade proporcional ao tamanho da população residente estimada pelo último censo demográfico anterior à data da pesquisa⁶.

No segundo estágio, as unidades de amostragem (setores censitários) são selecionadas, em cada município, com probabilidade proporcional ao tamanho, utilizando-se o número de unidades domiciliares por setor, obtido também pelo último censo demográfico. Em média, os setores censitários localizados nas zonas urbanas são compostos por 300 domicílios e os da zona rural, por 200.

Os domicílios são selecionados dentro de cada setor censitário escolhido no estágio anterior. Este procedimento é realizado com equi-probabilidade mediante uma seleção sistemática com partida aleatória, assemelhando-se em muito com uma amostra aleatória simples. São coletadas informações de todos os residentes em um domicílio selecionado.

Análise de dados amostrais complexos

A análise de dados provenientes de pesquisas por amostragem que incorpora

níveis de complexidade – estratificação, conglomeração (em um ou mais níveis) e probabilidades desiguais de seleção – é denominada análise de dados amostrais complexos. A incorporação do plano amostral na estimação de medidas descritivas populacionais permite a obtenção de estatísticas não viciadas com respeito ao desenho amostral.

É muito comum por parte de pesquisadores a utilização dos microdados (dados coletados) da pesquisa ignorando este tipo de informação. Ao calcular as estimativas de interesse, os pesquisadores assumem que os dados são provenientes de uma amostra aleatória simples (AAS), onde as observações são independentes e identicamente distribuídas (IID), ou utilizam os dados como se fossem informações coletadas a partir de um censo demográfico.

Os pacotes estatísticos, em geral, não possuem rotinas adequadas para incorporar as complexidades do plano amostral. As estimativas pontuais dos parâmetros são influenciadas pela ocorrência de pesos amostrais distintos, enquanto as estimativas de variância (precisão) dos estimadores dos parâmetros do modelo são influenciadas também pelos efeitos de estratificação e conglomeração⁷.

O estudo analítico (formulação de modelos) em uma pesquisa amostral é fundamentado na inferência associada à estimação pontual dos parâmetros de interesse, à precisão das estimativas pontuais e à construção de intervalos de confiança para as estimativas pontuais. O cálculo da variância das estimativas desempenha papel essencial na realização da inferência analítica, permitindo a avaliação da precisão das estimativas, bem como a construção de intervalos de confiança e a formulação de testes de hipóteses sobre os parâmetros dos modelos.

Um dos métodos para a avaliar o impacto da incorporação do plano amostral

⁴ Exceto a área rural da região Norte.

⁵ Municípios auto-representativos são aqueles com população maior do que um tamanho predefinido ou que fazem parte da região metropolitana.

⁶ Todos os municípios auto-representativos estão incluídos na amostra, só ocorrendo esta etapa de seleção para os estratos formados por municípios não auto-representativos.

⁷ Mais detalhes em Pessoa e Silva (1998) e Leite (2001).

sobre a precisão das estimativas foi desenvolvido por Kish, sendo denominado Efeito do Plano Amostral (EPA ou *Deff – Design Effect*), cuja expressão é dada por:

$$EPA = \frac{Var_{verd}(\hat{\beta})}{Var_{aas}(\hat{\beta})}$$

onde $Var_{verd}(\hat{\beta})$ é a variância estimada incorporando o plano amostral efetivamente utilizado e $Var_{aas}(\hat{\beta})$ é a variância estimada supondo o plano amostral igual a uma amostra aleatória simples.

O objetivo principal desta medida é a comparação entre planos amostrais no estágio de planejamento de pesquisas (Pessoa e Silva, 1998). Valores elevados do EPA destacam a importância da consideração do verdadeiro plano amostral efetivamente utilizado na estimação das variâncias associadas às estimativas dos parâmetros. As estimativas das variâncias baseadas em hipóteses de AAS subestimam os resultados corretos, conforme a descrição abaixo:

EPA < 1 → variância sob AAS superestimada;

EPA = 1 → não há diferença entre as estimativas de variância;

EPA > 1 → variância sob AAS subestimada.

Pessoa e Silva (1998) apontam uma certa dificuldade no cálculo da referida medida para fins analíticos. Skinner, Holt e Smith (1989) desenvolveram outra medida, contornando as dificuldades do EPA de Kish, denominada EPA ampliado (*Meff – Misspecification Effect*). Diferentemente da primeira, esta nova medida tem utilidade para fins analíticos e avalia a tendência de um estimador usual (consistente), calculado sob hipótese de IID, em subestimar ou superestimar a variância verdadeira do estimador pontual. O EPA ampliado é definido por:

$$EPA(\hat{\beta}; v_0) = \frac{V_{verd}(\hat{\beta})}{E_{verd}(v_0)}$$

onde $v_0 = \hat{V}_{IID}(\hat{\beta})$ é um estimador usual (consistente) da variância do estimador sob

a hipótese de observações IID; $V_{verd}(\hat{\beta})$ é a variância do estimador sobre o plano efetivamente utilizado; $E_{verd}(v_0)$ é a esperança do estimador usual sob o plano amostral efetivamente utilizado.

Quanto maior o valor do EPA e do EPA ampliado, menor será a probabilidade real de cobertura do intervalo de confiança para o parâmetro de interesse, caso o intervalo seja calculado sem considerar o plano amostral da pesquisa (cf. Pessoa e Silva, 1998). Faz-se necessária, então, a incorporação do plano amostral na inferência analítica quando os dados são provenientes de uma amostra probabilística.

A modelagem de dados provenientes de uma pesquisa amostral com níveis de complexidade sem a correta especificação do plano efetivamente utilizado pode apresentar problemas de superestimação ou subestimação dos parâmetros do modelo e das variâncias dos estimadores. No caso de modelagem paramétrica, a incorporação do plano amostral permite a obtenção de estimativas robustas. É preciso ter muito cuidado na produção de estatísticas provenientes de pesquisas com planos amostrais complexos, dado que não basta incorporar somente o fator de expansão (peso) de cada unidade amostral nos cálculos. Deve-se também considerar os efeitos de conglomeramento e estratificação da amostra em questão, pois estes afetam as variâncias das estimativas dos coeficientes, influenciando diretamente sobre a decisão a respeito da significância dos parâmetros do modelo. A maioria dos pacotes estatísticos considera os pesos amostrais como replicações, gerando estimativas para as variâncias muito subestimadas⁸.

Método de estimação de parâmetros incorporando o plano amostral

Seja i o índice que representa um elemento de uma população, tal que i contém (resume) as informações referentes ao estrato, à unidade primária de amostragem (UPA) e ao elemento em uma dada unidade

⁸ Detalhes em Pessoa e Silva (1998).

primária. Suponha que se observa o conjunto de variáveis aleatórias (y_i, \underline{x}_i) provenientes da população U , onde y_i é a variável resposta e \underline{x}_i é um vetor de características (variáveis explicativas) associadas a cada i , tal que $i \in U$. Adicionalmente, assume-se que y_1, \dots, y_N são IID, com função de densidade $f(y, \underline{\beta})$, onde $\underline{\beta}$ é um vetor de parâmetros desconhecidos de interesse. A função de verossimilhança e o logaritmo da função de verossimilhança populacional são descritos por:

$$L(\underline{\beta}; y_i, \underline{x}_i) = \prod_{i \in U} f(y_i, \underline{\beta})$$

$$l(\underline{\beta}; y_i, \underline{x}_i) = \sum_{i \in U} \log [f(y_i, \underline{\beta})]$$

sendo $l(\underline{\beta}; y_i, \underline{x}_i)$ o logaritmo da função de verossimilhança associado ao modelo, onde $\underline{\beta}$ é o vetor de parâmetros com dimensão $p \times 1$, \underline{x}_i é um vetor de dimensão $1 \times p$, definido para todo $i \in U$.

Então, para uma população finita, os parâmetros $\underline{\beta}$ são determinados através da solução de um sistema de equações definido por:

$$G(\underline{\beta}) = \sum_{i \in U} u_i(\underline{\beta}; y_i, \underline{x}_i)$$

onde $u_i = \frac{\partial l(\underline{\beta}; y_i, \underline{x}_i)}{\partial \underline{\beta}}$ é o vetor $p \times 1$ dos escores do elemento $i, \forall i \in U$.

O sistema acima é composto por p incógnitas que representam elementos do vetor de parâmetros $\underline{\beta}$. Igualando-se $G(\underline{\beta})$ a zero, a solução do sistema determina as p estimativas que compõem o vetor do estimador $\hat{\underline{\beta}}$, denominado estimador de máxima-verossimilhança para $\underline{\beta}$, quando y_i é conhecido $\forall i \in U$.

Numa pesquisa amostral não são coletadas informações sobre toda a população. Os elementos de uma amostra pertencem a um conjunto s selecionado da população U . Quando $i \in s$, $G(\underline{\beta}) = \sum_{i \in U} u_i(\underline{\beta}; y_i, \underline{x}_i)$

pode ser estimado empregando-se um estimador linear ponderado da forma $\hat{G}(\underline{\beta}) = \sum_{i \in s} w_i u_i(\underline{\beta}; y_i, \underline{x}_i)$, onde w_i são os pesos previamente definidos. O estimador de Pseudo Máxima-Verossimilhança de $\hat{\underline{\beta}}$ é a solução do sistema de equação, cujas equações são dadas por:

$$\hat{G}(\underline{\beta}) = \sum_{i \in s} w_i u_i(\underline{\beta}; y_i, \underline{x}_i) = 0$$

A partir da equação acima é possível estimar a variância assintótica de $\hat{\underline{\beta}}$ utilizando a matriz de primeira ordem da expansão da série de Taylor para o estimador de pseudo máxima-verossimilhança. Assim, o estimador da variância assintótica é dado por:

$$\hat{V}(\hat{\underline{\beta}}) = \left[\frac{\partial \hat{G}(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\underline{\beta} = \hat{\underline{\beta}}} \right]^{-1} \hat{V} \left[\sum_{i \in s} w_i u_i(\hat{\underline{\beta}}; y_i, \underline{x}_i) \right] \left[\frac{\partial \hat{G}(\underline{\beta})}{\partial \underline{\beta}} \Big|_{\underline{\beta} = \hat{\underline{\beta}}} \right]^{-1}$$

onde $\hat{V} \left[\sum_{i \in s} w_i u_i(\hat{\underline{\beta}}; y_i, \underline{x}_i) \right]$ é um estimador consistente para a variância do estimador do total populacional dos escores. Binder (1983) mostrou que a distribuição assintótica de $\hat{\underline{\beta}}$ é normal multivariada, fornecendo uma base para inferência sobre $\underline{\beta}$ sob amostras grandes, tal que $\hat{V}(\hat{\underline{\beta}})^{-1/2}(\hat{\underline{\beta}} - \underline{\beta}) \sim NM(0; I)$.

De acordo com Pessoa e Silva (1998), os pesos w_i devem ser tais que satisfaçam algumas condições, a saber: os estimadores devem ser assintoticamente normais, não viciados e com estimadores de variância consistentes. Estas condições são satisfeitas quando a probabilidade de inclusão na amostra da i -ésima unidade da população, $i = 1, 2, \dots, N$, é maior do que zero ($\pi_i > 0$) e, simultaneamente, a probabilidade de inclusão conjunta da i -ésima e j -ésima unidades da população, $i \neq j$, é também maior do que zero ($\pi_{ij} > 0$). Cabe ressaltar que os estimadores de pseudo máxima-verossimilhança não são únicos, pois existem diversas maneiras de definir os pesos w_i . Entretanto, utiliza-se usualmente o estimador de Horwitz-Thompson para totais, cujos pesos são dados por $w_i = \frac{1}{\pi_i}, \forall i \in s$.

O modelo multinomial logístico

O uso do modelo multinomial logístico é adequado para os casos em que a variável resposta é qualitativa e nominal com J possíveis categorias. Cabe ressaltar que neste caso não existe vantagem alocativa entre as J categorias, isto é, não existe ordenação entre as categorias (Powers e Xie, 2000).

Para a estimação é necessário definir uma das J possíveis categorias de resposta como base, isto é, as probabilidades das outras $J-1$ categorias serão estimadas com relação à categoria base. Qualquer categoria de resposta pode ser considerada base para a estimação dos parâmetros. De forma a facilitar a notação, a categoria de resposta (J) será sempre usada como base. A equação do modelo é definida como:

$$\ln\left(\frac{P(Y_j=j|X_j)}{P(Y_j=J|X_j)}\right) = \ln\left(\frac{P_j|X_j}{P_J|X_j}\right) = \ln\left(\frac{\theta_j}{\theta_J}\right) = \underline{x}_j \underline{\beta}_j \quad \forall j=1, \dots, J-1$$

onde $\underline{x}_j \underline{\beta}_j = \sum_{k=0}^K x_{jk} \beta_{jk}$.

A necessidade de ter uma categoria base, cujos coeficientes estimados serão iguais a zero, deve-se à questão de identificabilidade dada por:

$$\sum_{j=1}^J \theta_{ij} = 1, \Rightarrow \theta_{iJ} = 1 - \sum_{j=1}^{J-1} \theta_{ij} \quad \forall i.$$

A partir das equações acima é possível representar as probabilidades de cada categoria j como função das variáveis explicativas do modelo na forma:

$$\theta_{ij} = \frac{\exp(\underline{x}_i \underline{\beta}_j)}{\sum_{j=1}^J \exp(\underline{x}_i \underline{\beta}_j)}, \quad \forall j = 1, \dots, J$$

onde $\underline{\beta}_J = 0$.

Conclui-se, então, que para um modelo com K variáveis explicativas e J categorias são estimados $(K+1) \times (J-1)$ parâmetros. Nota-se, também, que o modelo logístico é um caso particular do modelo multinomial logístico quando $J = 2$.

Em geral, a interpretação dos parâmetros do modelo multinomial logístico é realizada através da análise das vantagens e razões de vantagens, bem como da especificação dos efeitos marginais das probabilidades para cada característica x_k de x .

No primeiro caso, a vantagem de pertencer à categoria j em relação a pertencer à categoria base J é dada por:

$$v_j = \frac{\theta_j}{\theta_J} = \exp(\underline{x}_j \underline{\beta}_j) \quad \forall j = 1, 2, \dots, J-1.$$

O logaritmo da vantagem é função linear de \underline{x}_j tal que:

$$\ln(v_j) = \ln\left(\frac{\theta_j}{\theta_J}\right) = \underline{x}_j \underline{\beta}_j \quad \forall j = 1, 2, \dots, J-1.$$

Quando a variável explicativa é representada por uma variável indicadora, ou medida em escala binária, a interpretação do efeito da variável na vantagem em favor da ocorrência do evento j é direta. Um coeficiente $\hat{\beta}_{jk}$ positivo indica que maior será a chance de ocorrer o evento na categoria j do que na categoria base J , enquanto um coeficiente negativo indica a ocorrência de relação oposta.

Quando o interesse é medir a vantagem da categoria j em relação a dois subgrupos diferentes de x_k , é necessário calcular a razão de vantagens. Seja θ_j a probabilidade de pertencer à categoria j e seja x_k uma variável contínua, então a razão de vantagens de $x_k = x_k^* + 1$ sobre $x_k = x_k^*$ é dada por:

$$\frac{\theta_j|x_k = x_k^* + 1}{\theta_j|x_k = x_k^*} = \frac{\exp((x_k^* + 1) \cdot \beta_{jk})}{\exp(x_k^* \cdot \beta_{jk})} = \exp(\beta_{jk}).$$

Generalizando-se estes conceitos, visando à comparação de qualquer par de categorias dentre as J categorias existentes, obtém-se:

$$\frac{v_j}{v_{j'}} = \frac{\theta_j/\theta_J}{\theta_{j'}/\theta_J} = \frac{\theta_j}{\theta_{j'}} = \exp[x_j(\beta_j - \beta_{j'})], \quad \forall j \neq j' = 1, \dots, J.$$

Logo, a diferença entre os coeficientes de x_k determina a direção da vantagem de pertencer à categoria j ou j' . Se a diferença é positiva, há maior vantagem de observar j do que j' .

Para avaliar o comportamento das probabilidades θ_j quando ocorrem mudanças em x_k é preciso calcular os efeitos marginais. Conforme observado, o modelo multinomial logístico não é um modelo linear, logo a influência da variável explicativa x_k sobre a probabilidade (θ_j) de pertencer à categoria de resposta j não é constante sobre todo o intervalo de variação de x_k . Segundo Powers e Xie (2000), o sinal do parâmetro (β_{jk}) associado à variável explicativa x_k não é necessariamente igual à mudança ocorrida na probabilidade θ_j , dada a condição de identificabilidade; ou seja, a probabilidade θ_j pode decrescer quando β_{jk} é positivo. Uma

das formas de verificar o efeito das variáveis explicativas sobre as probabilidades é através do efeito marginal calculado por (assumindo-se a categoria J como base):

$$\frac{\partial \theta_j}{\partial x_k} = \theta_j \left(\beta_j - \sum_{j=1}^J \theta_j \beta_j \right) \forall j = 1, \dots, J.$$

A partir da equação acima é possível estimar o impacto de *mudanças* marginais nas variáveis explicativas x_k sobre as probabilidades θ_j . Geralmente, os cálculos necessários para a solução da equação acima não são diretos quando a variável explicativa de interesse possui interação com alguma outra variável no modelo. Quando isto ocorre não é possível usar a referida equação para obter o efeito marginal.

Uma outra forma de mensurar o efeito marginal usando os coeficientes estimados no modelo é apresentada em Corseuil, Foguel e Santos (2001). A partir dos coeficientes estimados, calcula-se a probabilidade média de os indivíduos (θ_j) pertencerem à categoria de resposta j , através de mudanças no vetor de variáveis explicativas $x_i = (1, x_{i1}, \dots, x_{ik})$.

O procedimento consiste na criação de um novo vetor x_i^* , a partir da simulação de que todos os indivíduos na amostra possuem uma dotação K^* de uma determinada componente do vetor x_i . O segundo passo consiste em recalcular as probabilidades médias θ_j^* empregando-se x_i^* no lugar de x_i , utilizando como base a equação do modelo apresentada no início da seção. Em seguida, repete-se este procedimento atribuindo uma nova dotação K^{**} à mesma componente de x_i e calcula-se θ_j^{**} . A seqüência de probabilidades estimadas fornece uma boa estimativa do comportamento desta componente. É possível, também, modificar os valores de x_k mantendo-se todas as demais variáveis constantes em seus valores médios.

Análise do modelo multinomial logístico

A estimação dos parâmetros do modelo, incorporando o plano amostral, foi realizada utilizando-se o pacote estatístico STATA, versão 6.0. Como dito anteriormente, a variável resposta de interesse no estudo,

referente à situação ocupacional das crianças e dos adolescentes, é politômica, com quatro possíveis categorias de resposta, mutuamente exclusivas, definidas por:

- somente participa da força de trabalho (FT);
- somente estuda (E);
- participa da força de trabalho e estuda (FT/E);
- não realiza nenhuma atividade (NE/FF).

A escolha do elenco de (potenciais) variáveis explicativas para o modelo foi baseada na literatura disponível sobre o tema trabalho infantil. Corseuil, Foguel e Santos (2001) ressaltam que o nível educacional dos pais é um dos fatores determinantes para que crianças e adolescentes se afastem do mercado de trabalho. Quanto maior o nível educacional dos pais, maior a conscientização sobre a importância da educação. Adicionalmente, Patrinos e Psacharopoulos (1997) destacam a influência do tamanho e da composição do domicílio sobre a alocação ocupacional das crianças. Um estudo realizado por Hanusheck (1992) relaciona a qualidade da educação e o número total de crianças no domicílio. Estes estudos baseiam-se na hipótese de que os chefes dos domicílios possuem recursos limitados para o ensino dos filhos. Logo, quanto maior o número de filhos, menor será a disponibilidade de recursos para sua educação. Além disso, várias famílias não possuem recursos para a educação dos filhos ou dependem das crianças para aumentar o seu nível de renda. Ranjan (1999) associa o nível de renda familiar à alocação de tempo das crianças em sociedades carentes ou menos favorecidas.

Neste artigo, optou-se por utilizar uma variável composta pela renda domiciliar, excluindo-se os rendimentos dos menores de 18 anos residentes no domicílio, padronizada pelo total de moradores do domicílio. A inclusão de tal variável no modelo visa encontrar o efeito da necessidade da complementação da renda domiciliar sobre a inserção ocupacional de crianças e de adolescentes. Os argumentos apresentados acima conduzem à definição de

um conjunto de variáveis que, potencialmente, podem influenciar a escolha da criança ou do adolescente.

Devido à incorporação do plano amostral efetivamente utilizado pela PNAD/IBGE na estimação dos parâmetros, a variável raça não foi utilizada na análise dos dados. Em todos os testes estatísticos utilizados para a verificação do nível de significância das estimativas dos parâmetros observou-se que as estimativas encontradas para esta variável não eram significativas. Cabe destacar que se o mesmo modelo fosse estimado sem a incorporação do plano amostral, a mesma variável (raça) seria significativa a 1%, e com isso poderiam ser geradas análises equivocadas sobre o evento em estudo.

Sendo assim, o conjunto inicial de variáveis definidas para análise é dado por: *sexo*, *idade* e *educação (em anos de estudo) da criança*; *nível de instrução da mãe (em anos de estudo)*; *renda domiciliar per capita, exclusive a renda dos menores de 18 anos, expressa em salários mínimos*⁹; *razão entre o número de cômodos que servem de dormitórios e o total de cômodos do domicílio*; *taxa de dependência*¹⁰ *das crianças entre 0 e 4 anos (inclusive) residentes no domicílio*; *variável indicadora de boa condição de habitação*¹¹; *variáveis indicadoras de localização do domicílio, tais como região (Nordeste ou Sudeste), zona geográfica (rural ou urbana) e área geográfica (metropolitana/não-metropolitana).*

Foram consideradas também as diversas interações entre as variáveis (interações duplas), totalizando 66 variáveis explicativas, além do intercepto. Para avaliar a significância das diversas variáveis no modelo utilizou-se um teste estatístico da forma:

$$H_0: \beta_{jk} = 0 \quad H_a: \beta_{jk} \neq 0 \quad \forall_j = 1, \dots, J-1 \quad k = 0, 1, \dots, K$$

onde β_{jk} é o parâmetro associado à categoria j e à k -ésima variável explicativa.

Sob H_0 , a estatística de teste é definida como:

$$t^* = \frac{|\hat{\beta}_{jk} - 0|}{\sqrt{\hat{V}(\hat{\beta}_{jk})}} \sim t\text{-student}(u-L)$$

onde $V(\hat{\beta}_{jk})$ é a variância do estimador, j é uma dada categoria de escolha e $k = 0, 1, \dots, K$ são as variáveis explicativas, u é o número de unidades primárias de amostragem (3.288) e L é o número de estratos (213). Se $P(t^* > |t|) \leq 0,05$, há evidências para rejeitar H_0 .

O modelo ajustado permite analisar o comportamento da variável resposta condicionada a todas as variáveis explicativas simultaneamente. A partir da utilização do teste estatístico foi possível selecionar um modelo com 13 variáveis explicativas, além do intercepto, definido por: *anos de estudo (educa)*; *idade (idade)*; *indicador de sexo (hom: 1=homem)*; *taxa de cômodos (txcom)*; *anos de estudo da mãe (edu_mae)*; *renda domiciliar per capita em salários mínimos (rdpc_s)*; *taxa de dependência das crianças de 0 a 4 anos residentes no domicílio (txdep04)*; *indicador de boa condição habitacional (boa: 1 = boa condição)*; *indicador de região do domicílio (se: 1 = sudeste)*; *indicador de localização do domicílio (zona: 1 = urbana)*; *indicador de o domicílio estar em área metropolitana (metrop: 1 = área metropolitana)*; *interação entre ano de estudo e idade (A)*; e *interação entre domicílio localizado na região Sudeste e taxa de dependência das crianças de 0 a 4 anos residentes no domicílio (B)*.

O modelo proposto, que relaciona a probabilidade de ocorrência de uma categoria de situação ocupacional da criança ou adolescente θ_j com os fatores determinantes descritos, é dado por:

$$\ln \left(\frac{\theta_j}{\theta_J} \right) = \beta_{j0} + \beta_{j1}^{educa} + \beta_{j2}^{idade} + \beta_{j3}^{hom} + \beta_{j4}^{txcom} + \beta_{j5}^{edu_mae} + \beta_{j6}^{rdpc_s} + \beta_{j7}^{txdep04} + \beta_{j8}^{boa} + \beta_{j9}^{se} + \beta_{j10}^{zona} + \beta_{j11}^{metrop} + \beta_{j12}^A + \beta_{j13}^B$$

onde θ_j é a probabilidade de uma criança ou adolescente estar somente estudando.

⁹ O salário mínimo vigente em setembro de 1999 era de R\$ 136,00 (cento e trinta e seis reais).

¹⁰ Segundo Berquó (1991), "a razão de dependência é o quociente entre a 'população dependente' e a 'população potencialmente ativa' e, de acordo com o estudo citado, razões de dependência com valores elevados são vistas como refletindo desvantagens econômicas". O critério para escolha da população dependente é apenas cronológico, assumindo-se que indivíduos em determinada idade são inativos. Em geral, consideram-se como inativos os menores de 15 anos e maiores de 64 anos (Camarano, 1999).

¹¹ A condição habitacional é considerada boa se o domicílio possui água canalizada, escoamento sanitário e luz elétrica.

TABELA 1
Estimativas dos coeficientes e desvios-padrão

	Coeficiente	Estimativas Desvio-padrão	P-valor
FT/E			
Anos de estudo da criança ou do adolescente	-0,085	0,044	0,057
Idade da criança ou do adolescente	0,330	0,010	0,000
Sexo da criança ou do adolescente	1,002	0,047	0,000
Taxa de cômodos	0,880	0,146	0,000
Anos de estudo da mãe	-0,059	0,007	0,000
Renda domiciliar per capita em s.m.	-0,084	0,022	0,000
Taxa de dependência das crianças de 0 a 4 anos residentes no domicílio	0,196	0,169	0,247
Condições habitacionais boas	-0,212	0,064	0,001
Domicílio localizado na região Sudeste	-0,229	0,072	0,001
Domicílio localizado em zona urbana	-1,288	0,076	0,000
Domicílio localizado em área metropolitana	-0,555	0,058	0,000
Interação entre ano de estudo e idade da criança ou do adolescente	0,008	0,003	0,007
Interação localização na região Sudeste e taxa de dependência das crianças	-0,591	0,323	0,068
Intercepto	-5,580	0,152	0,000
FT			
Anos de estudo da criança ou do adolescente	-1,247	0,136	0,000
Idade da criança ou do adolescente	0,656	0,028	0,000
Sexo da criança ou do adolescente	1,108	0,066	0,000
Taxa de cômodos	0,486	0,219	0,027
Anos de estudo da mãe	-0,123	0,013	0,000
Renda domiciliar per capita em s.m.	-0,204	0,079	0,010
Taxa de dependência das crianças de 0 a 4 anos residentes no domicílio	1,294	0,253	0,000
Condições habitacionais boas	-0,394	0,100	0,000
Domicílio localizado na região Sudeste	0,760	0,105	0,000
Domicílio localizado em zona urbana	-0,982	0,096	0,000
Domicílio localizado em área metropolitana	-0,633	0,101	0,000
Interação entre ano de estudo e idade da criança ou do adolescente	0,064	0,008	0,000
Interação localização na região Sudeste e taxa de dependência das crianças	-0,419	0,400	0,295
Intercepto	-10,837	0,444	0,000
NE/FF			
Anos de estudo da criança ou do adolescente	-2,933	0,074	0,000
Idade da criança ou do adolescente	0,068	0,011	0,000
Sexo da criança ou do adolescente	-0,136	0,036	0,000
Taxa de cômodos	0,388	0,150	0,010
Anos de estudo da mãe	-0,082	0,008	0,000
Renda domiciliar per capita em s.m.	-0,209	0,039	0,000
Taxa de dependência das crianças de 0 a 4 anos residentes no domicílio	1,454	0,115	0,000
Condições habitacionais boas	-0,275	0,059	0,000
Domicílio localizado na região Sudeste	0,779	0,061	0,000
Domicílio localizado em zona urbana	-0,097	0,058	0,092
Domicílio localizado em área metropolitana	0,090	0,053	0,092
Interação entre ano de estudo e idade da criança ou do adolescente	0,171	0,004	0,000
Interação localização na região Sudeste e taxa de dependência das crianças	-0,452	0,181	0,012
Intercepto	-1,987	0,137	0,000

Fontes: PNAD/IBGE 1999 e cálculo dos autores.

TABELA 2
Efeito do plano amostral para as estimativas do modelo

	FT/E			FT			NE/FF		
	Var	DEFF	MEFF	Var	DEFF	MEFF	Var	DEFF	MEFF
Anos de estudo da criança ou do adolescente	0,002	1,6	1,5	0,019	1,4	1,7	0,005	1,2	1,5
Idade da criança ou do adolescente	0,000	1,8	1,4	0,001	1,7	2,4	0,000	1,8	2,6
Sexo da criança ou do adolescente	0,002	2,1	2,2	0,004	1,2	1,1	0,001	1,2	1,2
Taxa de cômodos	0,021	2,0	2,0	0,048	1,3	1,4	0,023	2,1	2,1
Anos de estudo da mãe	0,000	1,6	1,6	0,000	1,2	1,2	0,000	1,5	1,6
Renda domiciliar per capita em s.m.	0,000	1,6	2,0	0,006	1,4	2,6	0,002	1,0	2,4
Taxa de dependência das crianças de 0 a 4 anos residentes no domicílio	0,029	1,8	1,6	0,064	1,5	1,4	0,013	1,2	1,2
Condições habitacionais boas	0,004	2,3	2,2	0,010	1,5	1,5	0,003	1,7	1,8
Domicílio localizado na região Sudeste	0,005	3,4	3,3	0,011	2,0	2,0	0,004	1,7	1,8
Domicílio localizado em zona urbana	0,006	4,1	3,9	0,009	2,0	1,9	0,003	1,8	1,8
Domicílio localizado em área metropolitana	0,003	2,4	2,3	0,010	1,9	1,9	0,003	2,0	2,0
Interação entre ano de estudo e idade da criança ou do adolescente	0,000	1,5	1,5	0,000	1,4	1,6	0,000	1,3	1,4
Interação localização na região Sudeste e taxa de dependência das crianças	0,105	1,8	1,8	0,160	1,3	1,3	0,033	1,2	1,2
Intercepto	0,023	2,0	1,7	0,197	1,7	2,3	0,019	2,0	2,4

Fontes: PNAD/IBGE 1999 e cálculo dos autores.

Nota: Var - estimativa da variância dos estimadores dos parâmetros incorporando o plano amostral.

Para a estruturação do modelo final (reduzido) aplicou-se o procedimento de estimação que incorpora o desenho amostral da PNAD, no qual se definem apropriadamente os estratos, as unidades primárias de amostragem e os pesos amostrais. A Tabela 1 apresenta as estimativas obtidas para os parâmetros do modelo e respectivos desvios-padrão. As estimativas dos efeitos do uso do plano amostral EPA (*Deff*) e EPA ampliado (*Meff*), dispostas na Tabela 2, indicam que as estimativas de variância dos estimadores dos parâmetros obtidas desta forma são, em média, o dobro daquelas obtidas caso o plano amostral não fosse adequadamente especificado. Conclui-se que a incorporação do plano amostral é essencial para a obtenção de estimativas robustas e não viciadas.

De acordo com o modelo ajustado, a probabilidade de crianças e adolescentes entre 5 e 17 anos, residentes nas regiões Sudeste e Nordeste, estarem *somente participando da força de trabalho* (FT) é de 0,03, enquanto a probabilidade de estarem *somente estudando* (E) é de 0,77. A probabilidade de exercerem *ambas as atividades* (FT/E) foi estimada em 0,12 e a de *não estarem estudando, nem*

participando da força de trabalho, em 0,08 (NE/FF).

O efeito de cada variável explicativa (sem interações) é apresentado na Tabela 3, que contém as razões de vantagens e a variação na probabilidade obtida através do efeito marginal. Os efeitos de *idade, educação, residir no Sudeste e taxa de dependência* foram obtidos utilizando-se o procedimento descrito em Corseuil, Foguel e Santos (2001). De acordo com a Tabela 3, as probabilidades de uma criança ou adolescente somente participar da força de trabalho ou participar da força de trabalho e estudar simultaneamente são maiores do que a probabilidade de estar somente estudando quando: são do sexo masculino, residentes em domicílios com alta taxa de cômodos servindo de dormitório, e localizados nas áreas não-metropolitanas e rurais. Já a probabilidade de não realizar nenhuma atividade difere das duas outras categorias na questão do sexo e no fato de os domicílios estarem localizados em áreas metropolitanas.

A renda domiciliar e o nível de instrução da mãe são fatores determinantes para a decisão de permanecer somente estudando. Observa-se que o afastamento da

TABELA 3
Efeitos marginais e razões de vantagens para as variáveis explicativas sem considerar os efeitos de interação

	Coefficiente	Efeito marginal	R.Vantagem
FT/E			
<i>Sexo (homem)</i>	1,002	0,10	2,7
<i>Taxa de cômodos</i>	0,880	0,09	2,4
<i>Anos de estudo da mãe</i>	-0,059	-0,01	0,9
<i>Renda domiciliar per capita em s.m.</i>	-0,084	-0,01	0,9
<i>Condições habitacionais boas</i>	-0,212	-0,02	0,8
<i>Domicílio localizado em zona urbana</i>	-1,288	-0,13	0,3
<i>Domicílio localizado em área metropolitana</i>	-0,555	-0,06	0,6
E			
<i>Sexo (homem)</i>	0,000	-0,11	1,0
<i>Taxa de cômodos</i>	0,000	-0,12	1,0
<i>Anos de estudo da mãe</i>	0,000	0,01	1,0
<i>Renda domiciliar per capita em s.m.</i>	0,000	0,03	1,0
<i>Condições habitacionais boas</i>	0,000	0,05	1,0
<i>Domicílio localizado em zona urbana</i>	0,000	0,15	1,0
<i>Domicílio localizado em área metropolitana</i>	0,000	0,06	1,0
FT			
<i>Sexo (homem)</i>	1,108	0,03	3,0
<i>Taxa de cômodos</i>	0,486	0,01	1,6
<i>Anos de estudo da mãe</i>	-0,123	0,00	0,9
<i>Renda domiciliar per capita em s.m.</i>	-0,204	-0,01	0,8
<i>Condições habitacionais boas</i>	-0,394	-0,01	0,7
<i>Domicílio localizado em zona urbana</i>	-0,982	-0,02	0,4
<i>Domicílio localizado em área metropolitana</i>	-0,633	-0,02	0,5
NE/FF			
<i>Sexo (homem)</i>	-0,136	-0,02	0,9
<i>Taxa de cômodos</i>	0,388	0,02	1,5
<i>Anos de estudo da mãe</i>	-0,082	-0,01	0,9
<i>Renda domiciliar per capita em s.m.</i>	-0,209	-0,01	0,8
<i>Condições habitacionais boas</i>	-0,275	-0,02	0,8
<i>Domicílio localizado em zona urbana</i>	-0,097	0,01	0,9
<i>Domicílio localizado em área metropolitana</i>	0,090	0,01	1,1

Fontes: PNAD/IBGE 1999 e cálculo dos autores.

criança e do adolescente desta atividade está relacionado com a incidência de pobreza nas famílias, evidenciada pela sua localização geográfica, pelas condições habitacionais, nível de renda domiciliar *per capita* e nível educacional das mães.

Continuando na Tabela 3, observa-se que, analisando os resultados a partir do efeito marginal, o impacto da variável sexo na probabilidade de ocorrência de uma dada escolha indica que as crianças e adolescentes do sexo masculino têm maiores chances de estarem inseridos no mercado de trabalho, pois o efeito marginal indica um aumento na probabilidade de so-

mente trabalhar (+0.03 pontos na probabilidade) e realizar ambas as atividades (+0,10). Adicionalmente, as probabilidades de somente estudar (-0,11) ou não realizar nenhuma atividade (-0,02) são menores entre aqueles do sexo masculino, revelando que as meninas têm maiores chances de pertencer a esta categoria.

Observa-se, também, que o nível de instrução da mãe e o nível de renda domiciliar *per capita* do domicílio em que as crianças e adolescentes residem influenciam as probabilidades estimadas na categoria *somente estudar*. O acréscimo de 1 ano de estudo na escolaridade materna

implica uma variação de 0,01 ponto na probabilidade de o filho estar somente estudando, enquanto a probabilidade de pertencer a todas as outras categorias diminui. Tal efeito torna-se mais evidente ao considerarmos a variação de 1 salário mínimo na renda domiciliar *per capita*, pois, neste caso, a probabilidade de estar só estudando aumenta em 0,03.

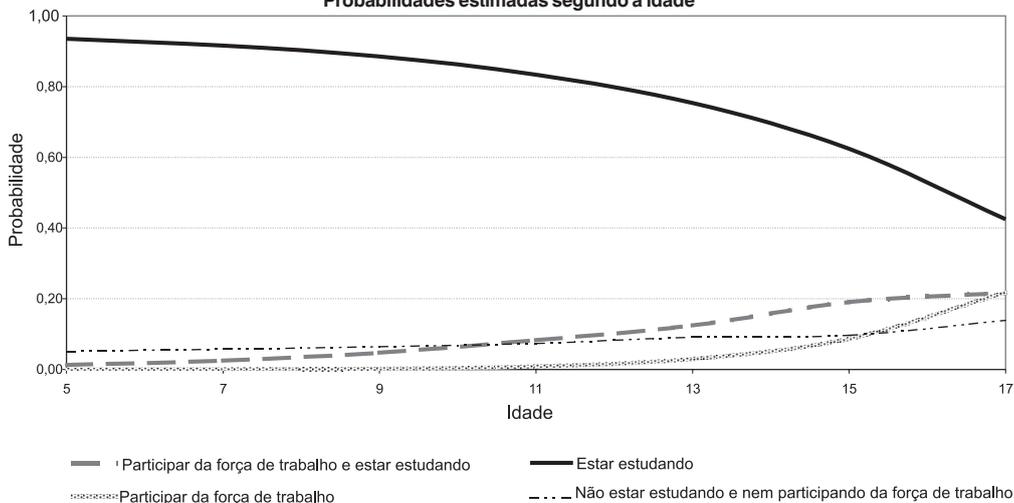
As crianças e os adolescentes residentes em domicílios localizados nas zonas urbanas têm maior probabilidade de estarem somente estudando, enquanto aqueles que residem em áreas rurais têm maiores probabilidades de pertencer à categoria *participa da força de trabalho e estuda*. Ao comparar as áreas metropolitanas e não-metropolitanas, verifica-se que a probabilidade de estar realizando ambas as atividades aumenta quando a criança reside em áreas não-metropolitanas, mas a probabilidade de não realizar nenhuma atividade aumenta caso o domicílio seja localizado em áreas metropolitanas.

Para a variável *idade*, conclui-se que quanto maior for a idade da criança ou do adolescente, maior será a probabilidade de estar somente trabalhando ou trabalhando e estudando ao mesmo tempo (Gráfico 1). De fato, tal comportamento não surpreende, pois o mercado de trabalho torna-se mais

atrativo entre os mais velhos. Além disso, supõe-se que quanto maior a idade da criança, maior é seu grau de instrução, a menos que exista atraso escolar. Não é possível afirmar que existe relação causal entre a entrada no mercado de trabalho e o atraso escolar, mas a introdução no modelo do efeito de interação entre as variáveis – *idade e anos de estudo da criança* – objetivou avaliar como a entrada no mercado de trabalho pode estar condicionada, principalmente, à idade (da criança ou do adolescente) e não especificamente ao atraso escolar. Observa-se, no Gráfico 2, que a probabilidade de uma criança ou adolescente com quatro anos de estudo pertencer à categoria *participa da força de trabalho e estuda* varia entre 0,10 e 0,24 dependendo da idade, isto é, uma criança de 11 anos (sem atraso escolar) tem uma probabilidade de estar realizando ambas as atividades 0,14 pontos menor do que um adolescente com 17 anos de idade (6 anos de atraso escolar). Este mesmo comportamento é observado quando são consideradas crianças e adolescentes que estão apenas participando da força de trabalho.

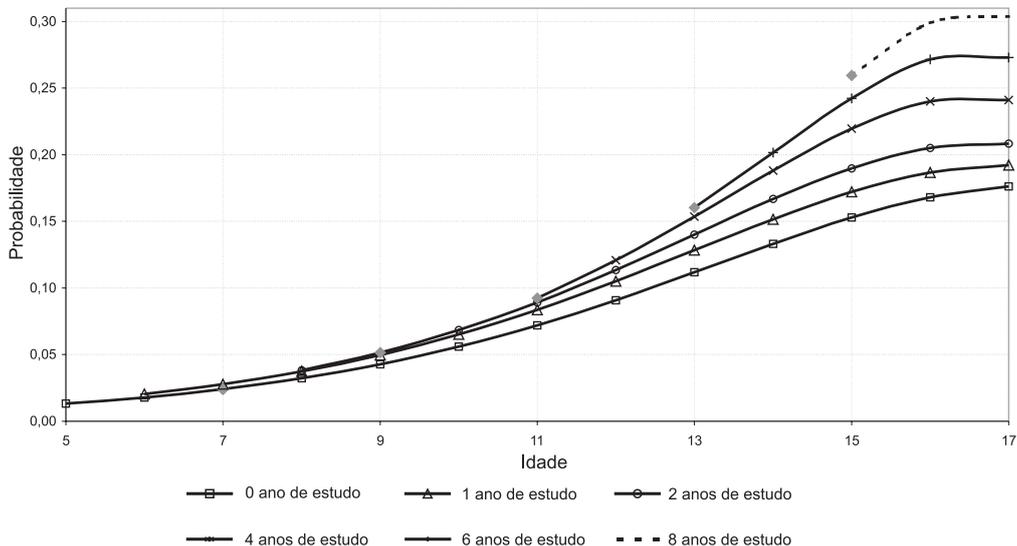
Dando continuidade à análise dos resultados, na Tabela 4 observa-se o efeito do número de pessoas residentes no domicílio no qual reside, pelo menos, uma

GRÁFICO 1
Probabilidades estimadas segundo a idade



Fontes: PNAD/IBGE 1999 e cálculo dos autores.

GRÁFICO 2
Probabilidades estimadas de pertencer à categoria Participa da Força de Trabalho e Estuda segundo a idade e o nível educacional



Fontes: PNAD/IBGE 1999 e cálculo dos autores.

TABELA 4
Probabilidades estimadas das categorias de interesse segundo a taxa de dependência infantil, sexo da criança ou do adolescente e localização do domicílio

Taxa de dependência	Probabilidades Estimadas								
	FT/E	Sexo masculino				Sexo feminino			
		E	FT	NE/FF	FT/E	E	FT	NE/FF	
0,0	0,15	0,74	0,04	0,06	0,08	0,83	0,02	0,08	
0,2	0,15	0,73	0,04	0,08	0,08	0,81	0,02	0,09	
0,4	0,14	0,71	0,05	0,10	0,07	0,79	0,02	0,11	
0,6	0,13	0,69	0,06	0,11	0,07	0,77	0,03	0,13	
0,8	0,13	0,67	0,07	0,14	0,07	0,75	0,03	0,16	
1,0	0,12	0,65	0,08	0,16	0,06	0,72	0,03	0,18	
2,0	0,08	0,50	0,12	0,30	0,04	0,56	0,05	0,34	

Taxa de dependência	Probabilidades Estimadas								
	FT/E	Sudeste				Nordeste			
		E	FT	NE/FF	FT/E	E	FT	NE/FF	
0,0	0,08	0,82	0,02	0,07	0,16	0,74	0,03	0,07	
0,2	0,07	0,81	0,03	0,08	0,16	0,71	0,04	0,09	
0,4	0,07	0,80	0,03	0,10	0,16	0,69	0,05	0,11	
0,6	0,06	0,79	0,03	0,11	0,15	0,66	0,05	0,13	
0,8	0,06	0,77	0,04	0,13	0,15	0,63	0,06	0,16	
1,0	0,05	0,76	0,04	0,15	0,14	0,59	0,07	0,20	
2,0	0,03	0,64	0,07	0,27	0,10	0,39	0,12	0,39	

Região	Probabilidades Estimadas			
	FT/E	E	FT	NE/FF
Nordeste	0,13	0,79	0,02	0,06
Sudeste	0,10	0,76	0,04	0,11

Fontes: PNAD/IBGE 1999 e cálculo dos autores.

criança ou adolescente entre 5 e 17 anos de idade. Note-se que quanto maior o valor da variável referente à taxa de cômodos, maior a chance de a criança ou o adolescente migrar da categoria *somente estuda* para a categoria *participa da força de trabalho e estuda*.

Adicionalmente, quanto maior a taxa de dependência, indicando a presença de várias crianças menores de 4 anos residindo no domicílio, maior é a chance de ocorrer migração da categoria *somente estuda* em direção às categorias *não realiza nenhuma atividade e somente trabalha* (Tabela 4). Este comportamento da variável taxa de dependência indica a importância deste fator para a caracterização dos componentes da categoria *não realiza nenhuma atividade*. O deslocamento obtido, dado um aumento da taxa de dependência, entre as crianças e os adolescentes que somente estudam ou realizam ambas as atividades para as outras duas categorias ocorre com intensidade diferente entre aqueles do sexo masculino e do sexo feminino. Os homens tendem a somente trabalhar, ao passo que as mulheres tendem a abandonar as atividades. Comparando-se as duas regiões geográficas, percebe-se que residir na região Nordeste aumenta a probabilidade de pertencer à categoria *participa da força de trabalho e estuda*, enquanto residir na região Sudeste aumenta as probabilidades de somente trabalhar ou não realizar nenhuma atividade (Tabela 4).

Conclusão

Dentre as 28,5 milhões de crianças e adolescentes com idade entre 5 e 17 anos residentes nas regiões Sudeste e Nordeste, estimou-se que 77% *só estudam* (E), enquanto 3% *somente participam da força de trabalho* (FT), 12% *participam da força e estudam* (FT/E) e 8% *não estudam e não participam da força de trabalho* (NE/FF).

As crianças e adolescentes do sexo masculino têm maiores chances de participar do mercado de trabalho, ao passo que as do sexo feminino têm maiores chances de não estudar e não participar da força de trabalho.

A participação no mercado de trabalho é crescente com a idade, mas o aumento das chances de o adolescente trabalhar ou não exercer nenhuma das duas atividades é maior entre os que possuem algum atraso escolar. Não há, entretanto, indícios de causalidade entre atraso escolar e a participação no mercado de trabalho. Destaca-se o fato de que, devido ao atraso escolar, os mais velhos exerçam alguma atividade econômica em paralelo à escola, ou mesmo abandonem a escola ao ficarem desestimulados com os estudos.

O papel da educação das mães, bem como o da renda domiciliar *per capita*, é fundamental para que a criança ou o adolescente se dedique somente aos estudos. Quanto melhor for o nível de instrução da mãe, ou menor a condição de pobreza do domicílio, maior será a probabilidade de o filho só estudar. Vale lembrar que a renda domiciliar *per capita* baixa não é condição para a entrada no mercado de trabalho, mas famílias com renda domiciliar *per capita* mais alta tendem a manter as crianças ou adolescentes fora do mercado de trabalho. O resultado encontrado para a influência da educação da mãe ratifica os resultados de Corseiul, Foguel e Santos (2001) e de Leme e Wajnman (2000). Em contrapartida, Corseiul, Foguel e Santos (2001) não observam nenhum efeito da condição de pobreza da família sobre a decisão do jovem em trabalhar ou estudar. Entretanto, observou-se neste artigo efeito semelhante ao encontrado por Ranjan (1999).

Com relação à taxa de dependência das crianças de 0 a 4 anos residentes no domicílio, há evidências de que quanto maior a taxa, maior a probabilidade de a criança ou o adolescente do sexo feminino não participar da força de trabalho e não estudar para poder cuidar das crianças mais novas da família. As crianças e adolescentes do sexo masculino tendem a participar da força de trabalho quanto maior for a taxa de dependência. Estes resultados confirmam a hipótese de Hanusheck (1992) de que um número maior de filhos prejudica a educação oferecida pelos pais, bem como a tese de Leme e Wajnman (2000) sobre a escolha das mulheres em não realizar

nenhuma das atividades em domicílios com muitas crianças.

As crianças e os adolescentes residentes em domicílios localizados na região Sudeste têm maior probabilidade de não participar da força de trabalho e não estudar, assim como maior probabilidade de somente participar do mercado de trabalho. Este fato indica que no Sudeste as adversidades criam situações mais propícias ao abandono escolar, apesar do menor desenvolvimento e maior condição de pobreza do Nordeste, onde, em geral, as crianças trabalham e estudam ao mesmo tempo. Quando o domicílio está localizado em zonas rurais, maiores são as chances de as crianças ou os adolescentes pertencerem à força de trabalho. Nas áreas metropolitanas observa-se um pequeno aumento na chance de a criança ou o adolescente não participar da força de trabalho e não estudar.

A partir desta análise, conclui-se que as crianças e os adolescentes residentes nas regiões Sudeste e Nordeste do Brasil têm maior dificuldade de estar só estudando devido à condição de pobreza observada em suas famílias. Estas crianças e adolescentes comprometem sua vida futura quando não pertencem à categoria *só estuda*. Ao exercerem alguma atividade econômica ou mesmo abandonarem a escola, estas crianças ou adolescentes convivem com a

possibilidade de permanecerem na condição de pobreza vivida pelos seus pais.

As chances de a criança ou o adolescente estarem inseridos no mercado de trabalho são: crescentes com a idade; mais altas entre aqueles do sexo masculino; dependentes do nível de instrução da mãe e da condição de pobreza; mais altas quanto menor o grau de urbanização da área em que residem; e crescentes com a taxa de dependência. Por outro lado, a chance de não exercer nenhuma atividade difere em dois pontos da conclusão anterior: é mais alta entre as crianças e adolescentes do sexo feminino, além de mais alta quanto maior o grau de urbanização da área em que residem.

Finalizando, destaca-se a utilização correta na análise dos dados de uma pesquisa amostral com níveis de complexidade. Os resultados obtidos por outros autores (Corseuil, Foguel e Santos, 2001; Leme e Wajnman, 2000), em estudos citados, não diferem muito dos resultados apresentados neste artigo, exceto pela estimação de parâmetros viciados tanto para as estimativas pontuais dos parâmetros como para a variância dos parâmetros do modelo. Estes estudos não utilizam o fator de expansão da amostra, considerando a mesma como se fosse um censo demográfico¹², além de não realizarem testes estatísticos robustos para a seleção de variáveis a serem utilizadas no modelo.

Referências bibliográficas

AGRESTI, A. **Categorical data analysis**. Nova York: John Wiley, 1990.

BARROS, R., HENRIQUES, R. e MENDONÇA, R. Pelo fim das décadas perdidas: educação e desenvolvimento sustentado no Brasil. In: HENRIQUES, R. (ed.). **Desigualdade e pobreza no Brasil**. Rio de Janeiro: IPEA, 2000.

BERQUÓ, E. Fatores estáticos e dinâmicos (mortalidade e fecundidade). In: SANTOS, J., LEVY, M. S. e SZMRECSÁNYI, T. **Dinâmica da população: teoria, métodos e técnicas de análise**. São Paulo: T. A. Queiroz, 1991.

BINDER, D. A. On the variances of asymptotically normal estimators from complex survey. **International Statistical Review**, v. 51, 1983. p. 279-292.

BRASIL. **Constituição (1988)**. Constituição da República Federativa do Brasil. Brasília, DF: Senado, 1988.

_____. **Consolidação das Leis de Trabalho**. Decreto-lei n. 5.452, de 1º de maio de 1943. Aprova a Consolidação das Leis do Trabalho. Lex-Coletânea de Legislação. Edição federal. São Paulo, v. 7, 1943. Suplemento.

¹² Veja detalhes em Leite (2001).

CAMARANO, A. e KANSO, S. Idosos brasileiros: que dependência é essa? In: CAMARANO, A. A. (org.). **Muito além dos 60: os novos idosos brasileiros**. Rio de Janeiro: IPEA, 1999.

CORSEIUL, C.H., FOGUEL, M. e SANTOS, D. Decisões críticas em idades críticas: a escolha dos jovens entre estudo e trabalho em quatro países da América Latina. **Texto para discussão**, IPEA, n. 797, 2001.

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

LEITE, P.G. **Análise da situação ocupacional de crianças e adolescentes nas regiões Sudeste e Nordeste do Brasil utilizando informações da PNAD 1999**. Dissertação de Mestrado, Escola Nacional de Ciências Estatísticas/IBGE, Rio de Janeiro, 2001.

LEME, M. e WAJNMAN, S. **A alocação do tempo dos adolescentes brasileiros entre o trabalho e a escola**. Trabalho apresentado no XII Encontro Nacional de Estudos Populacionais, Caxambu, MG, 2000.

MISSE, M. As ligações perigosas. Mercado informal legal, narcotráfico e violência no Rio. **Revista Semestral de Ciências Sociais e Educação**, IEC, Ano II, n. 1, 1997.

PATRINOS, H. e PSACHAROPOULOS, G. Family size, schooling and child labor in Peru –

an empirical analysis. **Journal of Population Economics**, v. 10, 1997. p. 387-405.

PESSOA, D. e SILVA, P. **Análise de dados amostrais complexos**. Trabalho apresentado no 13º Simpósio Nacional de Probabilidade e Estatística, promovido pela Associação Brasileira de Estatística, 1998.

POWERS, D. e XIE, Y. **Statistical methods for categorical analysis**. Academic Press, 2000.

RANJAN, P. An economic analysis of child labor. **Economic Letters**, v. 64, 1999. p. 99-105.

SILVA, N. e KASSOUF, A. O trabalho e a escolaridade dos jovens brasileiros. **Qualidade de vida**, Centro de Estudos Avançados em Economia Aplicada, Departamento de Economia, Administração e Sociologia, USP, Ano 3, n. 23, abril 2001.

SKINNER, C.J., HOLT, D. e SMITH, T.M.F. **Analysis of complex surveys**. Chichester: John Wiley, 1989.

SOUTO DE OLIVEIRA, J. Quando seremos apenas crianças. **Revista Democracia/PG**, n. 101, abril/maio 1994. p.12-16.

_____. **Juventude pobre: o desafio da integração**. Tese de Doutorado, Universidade do Estado do Rio de Janeiro, Instituto de Medicina Social, 1999.

STATA. **Stata user's guide**. Release 6. College Station, Texas: Stata Press, 1999.

Abstract

Using data from the 1999 Brazilian National Household Sample Survey (PNAD/IBGE), this article paper investigates the individual and family characteristics of children and adolescents aged 5 to 17 who live in the southeastern and the northeastern regions of Brazil regarding occupational choices. The options considered were: working only, studying only, both working studying, and neither. All estimations necessary for the study incorporated the PNAD sample design. We found that the occupational behavior of the children/adolescents is related to their living conditions and level of poverty. The less educated the mothers, the more likely are the children/adolescents to drop out of school. Children living in the northeastern region have a higher probability of both studying and working simultaneously, whereas in the southeastern region they are more likely to be unemployed work.

Enviado para publicação em 25/11/2002.