



Métodos para estimação da subenumeração do grupo etário de 0 a 4 anos no Censo Demográfico brasileiro de 2000

Reinaldo Onofre Santos*
Guilherme Quaresma Gonçalves**

O objetivo deste trabalho é discutir a aplicação de técnicas que contribuam para identificar e corrigir a subenumeração censitária para o grupo etário de 0 a 4 anos, utilizando, como população base, aquela enumerada no Censo Demográfico de 2000. Frequentemente, esse grupo possui baixa enumeração, se comparada àquela de faixas etárias da população adulta. Foram aplicadas cinco técnicas para a estimação do grupo de 0 a 4 anos e os resultados indicam que, não obstante o fato de técnicas já comumente usadas oferecerem resultados satisfatórios, formas alternativas têm grande potencial a ser explorado.

Palavras-chave: Censos demográficos. Subenumeração. Técnicas de análise demográfica.

* Centro de Desenvolvimento e Planejamento Regional (Cedeplar), Universidade Federal de Minas Gerais (UFMG), Belo Horizontes-MG, Brasil (reinaldosantos80@gmail.com; <https://orcid.org/0000-0001-6762-9100>).

** Centro de Desenvolvimento e Planejamento Regional (Cedeplar), Universidade Federal de Minas Gerais (UFMG), Belo Horizontes-MG, Brasil (gui.quaresma89@gmail.com; <https://orcid.org/0000-0002-0507-9956>).

Introdução

O presente trabalho traz uma discussão acerca da correção da enumeração censitária para o grupo etário de 0 a 4 anos. Geralmente, esse grupo sofre o efeito de duas fontes de erro: falhas na cobertura censitária; e omissão de indivíduos, principalmente crianças, no momento da declaração. Em linhas gerais, a subenumeração é um fator limitante para as projeções populacionais e a construção de estimativas confiáveis, uma vez que altera o valor da população base e do denominador de diversas taxas importantes.

A pesquisa de avaliação do Censo Demográfico de 2000 (OLIVEIRA et al., 2003) indica uma variação grande entre as unidades da federação quanto à qualidade da enumeração desse grupo etário. A estimativa do erro por omissão, do grupo de 0 a 1 ano, ficou entre 1,2% (Tocantins) e 11,9% (Pará). Já para o grupo de 1 a 4 anos, a menor subenumeração teria sido em Minas Gerais (0,6%) e a maior em Roraima (9,5%). A aplicação de métodos¹ que apontam, mesmo que parcialmente, a qualidade de enumeração censitária tem maior eficiência para a identificação do problema na população adulta, o que demanda alternativas para os grupos etários mais jovens e aqueles mais avançados da população.

No Manual III das Nações Unidas existe a recomendação de utilizar dados sobre os nascimentos ocorridos durante os cinco anos anteriores ao censo e aplicar uma razão de sobrevivência, do período, a esse grupo (UNITED NATIONS, 1956). Assim, ter-se-ia o total de crianças de 0 a 4 anos, sobreviventes, no ano do censo. Entretanto, essa alternativa apresenta algumas limitações. Internamente ao grupo, a subenumeração esperada é maior no subgrupo com menos de um ano de idade. Além disso, essa alternativa depende tanto da boa qualidade do registro de nascimentos como da função de mortalidade, esta última obtida de forma indireta.

No Brasil, o Instituto Brasileiro de Geografia e Estatística (IBGE) reconstituiu a população base das projeções, revisão de 2013,² a partir da média aritmética dos valores determinados por dois métodos: com base no Registro Civil de nascimentos anteriores ao Censo Demográfico de 2000, aplicando uma razão de sobrevivência; e a retroprojeção da população entre 10 e 14 anos, de 2010 (IBGE, 2013).

Neste trabalho, busca-se discutir algumas técnicas que, potencialmente, poderiam contribuir para aumentar a qualidade da cobertura do grupo de 0 a 4 anos da população base das projeções, lançando mão de cinco técnicas: estimação a partir do Registro Civil; retroprojeção do grupo de 10 a 14 anos; retroprojeção de nascimentos; relação entre a razão criança-mulher e a taxa de fecundidade total; e estimação da taxa bruta de natalidade por meio de duas estruturas etárias.

¹ Ver United Nations (2002) para entendimento do método *General Growth Balance* e sua contribuição para ajustar a enumeração média entre dois censos.

² Não se analisou a revisão 2018 porque esta, ao contrário da revisão de 2013, utiliza como população base o Censo Demográfico de 2010, não o de 2000, objeto de análise deste trabalho (IBGE, 2018).

A necessidade de utilizar dois censos para a retroprojeção do grupo de 10 a 14 anos (de acordo com o segundo método indicado no parágrafo anterior) impõe que o exercício de avaliação da enumeração do grupo etário de 0 a 4 anos seja possível apenas para o Censo Demográfico de 2000, uma vez que não há nenhum levantamento censitário após 2010, data do último censo demográfico.

Materiais e métodos

Para a aplicação dos métodos, foram utilizados os dados do Registro Civil de nascimentos, entre 1995 e 2000, as funções de sobrevivência retiradas das Tabelas de Vidas da revisão de 2013 das projeções do IBGE (IBGE, 2013), bem como os Censos Demográficos de 1991, 2000 e 2010.

Optou-se por aplicar as técnicas de correção do grupo de 0 a 4 anos para o total do Brasil, especificamente em 2000, pelos seguintes motivos: estoque do Registro Civil suficientemente atualizado no que se refere aos registros tardios de nascimento; o fato de o Censo Demográfico de 2000 estar entre os de 1991 e 2010, facilitando, assim, a retroprojeção do grupo de 10 a 14 anos, bem como a estimação das funções necessárias, para o quinquênio 1995-2000; e a menor influência das migrações internacionais sobre a totalidade do país, se comparada com o impacto que os fluxos externos e internos provocam nas unidades da federação.

A seguir são descritas cada uma das técnicas empregadas, bem como a fonte de dados utilizada.

Estimativa por meio do Registro Civil

A estimação no número de crianças entre 0 e 4 anos, em 2000, a partir do Registro Civil de nascimentos, segue rigorosamente as orientações do Manual III das Nações Unidas (UNITED NATIONS, 1956). Nesse caso,

$${}_5N_0^t = \frac{{}_5L_0^{t-5,t}}{{}_5l_0^{t-5,t}} \sum_{t-5}^t B = B^{t-5,t} * \frac{{}_5L_0^{t-5,t}}{{}_5l_0^{t-5,t}} \quad (1)$$

Onde:

${}_5N_0^t$ é o número de crianças de 0 a 4 anos, no ano t . Nesse caso, t refere-se a 2000;

${}_5L_0^{t-5,t}$ é o número total de anos vividos de uma população estacionária entre 0 e 5 anos, dados os níveis de mortalidade do quinquênio $t-a$ a t . Pode ser entendido também como o número médio de anos vividos por indivíduos nascidos entre $t-5$ e t durante os primeiros cinco anos desde o nascimento;

$l_0^{t-5,t}$ é a raiz da tabela de vida média do quinquênio anterior ao censo ($t-5$ a t);

$\sum_{t-5}^t B$ ou $B^{t-5,t}$ correspondem ao número total de nascidos vivos durante o período $t-5$ a t .

Além da dependência de uma tabela de vida estimada de forma indireta, esse método possui, como limitação, o fato de que parte do registro de nascimentos é feita de forma tardia, dificultando o uso dessa técnica imediatamente após o ano do censo. Quando há

inclusão do registro tardio, parte-se do pressuposto de que não existe seletividade da mortalidade dentre aqueles que foram registrados *a posteriori*. Se isso não se sustenta, haverá um erro, por falta, no número de pessoas no grupo etário.

Retroprojeção da população de 10 a 14 anos, utilizando dois censos

Uma alternativa para a estimação do primeiro grupo é a retroprojeção da população a partir de um segundo censo, no caso brasileiro, com intervalo decenal. Com auxílio de duas tabelas de sobrevivência quinquenais, bem como a estrutura etária do segundo censo, tem-se:

$${}_5N_0^t = {}_5N_0^{t+10} \frac{{}_5L_0^{t,t+5}}{{}_5L_5^{t,t+5}} * \frac{{}_5L_5^{t+5,t+10}}{{}_5L_{10}^{t+5,t+10}} \quad (2)$$

Onde: ${}_5L_0^{t,t+5}$, ${}_5L_5^{t,t+5}$, ${}_5L_5^{t+5,t+10}$ e ${}_5L_{10}^{t+5,t+10}$ descrevem o número total de anos vividos de uma população estacionária entre 0 a 4 e 5 a 9 anos, dados os níveis de mortalidade no primeiro (t , $t+5$) e segundo ($t+5$, $t+10$) quinquênios posteriores ao primeiro censo demográfico.³

As limitações correspondem à já esperada dependência da qualidade das estimativas da tabela de sobrevivência, ao menos até dez anos de idade, que devem representar o nível e a estrutura da mortalidade do decênio posterior ao censo de referência, assim como o fato de que a população estimada de 0 a 4 anos dependerá da exatidão da população enumerada de 10 a 14 anos no segundo censo, que pode estar subenumerada, ou, mais comumente, sobrenumerada, devido à atração do dígito 0.

Retroprojeção de nascimentos

A retroprojeção dos nascimentos consiste em estimar os nascimentos a partir de uma função média de fecundidade do quinquênio anterior ao censo (1995 a 2000), das razões de sobrevivência para o mesmo quinquênio e da distribuição etária de mulheres em idade reprodutiva no ano do censo. Após o cálculo dos nascimentos, aplica-se a função de sobrevivência do primeiro grupo etário.

Sabe-se que:

$$B^{t-5,t} = 5 \sum_{15}^{54} {}_nF_x^{t-5,t} * {}_nW_x^{t-2,5} \quad (3)$$

Onde:

$B^{t-5,t}$ corresponde ao número total de nascidos vivos durante o período $t-5$ a t ;

${}_nW_x^{t-2,5}$ é o número médio anual de mulheres entre x e $x+5$ anos, no quinquênio anterior ao ano do censo ($t-5$ a t), ou ainda o número de mulheres em $t-2,5$;

${}_nF_x^{t-5,t}$ expressa as taxas específicas de fecundidade dos grupos quinquenais de idade, dentro do quinquênio anterior ao ano do censo ($t-5$ a t).

³ Na ausência de tabelas de sobrevivência quinquenais, pode-se usar uma que espelhe a experiência média de mortalidade do decênio intercensitário (UNITED NATIONS, 2002).

Se supusermos que o número de mulheres-ano no grupo etário entre x e $x+5$ anos, entre os anos $t-5$ e t , seja a média aritmética entre o número de mulheres, de x e $x+5$ anos, nos tempos $t-5$ e t , temos que:

$${}_nW_x^{t-2,5} = 0,5({}_nW_x^{t-5} + {}_nW_x^t) \quad (4)$$

Contudo, considerando que as mulheres com idade entre x e $x+n$ no tempo $t-5$ correspondem à coorte de mulheres com idade entre $x+n$ e $x+2n$ no tempo t , aplicado o inverso da razão de sobrevivência no período, reconstituindo as mulheres que faleceram no período, tem-se:⁴

$${}_nW_x^{t-2,5} = 0,5 \left({}_nW_x^t + {}_nW_{x+n}^t \frac{{}_nL_x^{t-5,t}}{{}_nL_{x+n}^t} \right) \quad (5)$$

Substituindo a equação (5) na equação (3):

$$B^{t-5,t} = 5 \sum_{15}^{54} {}_nF_x^{t-5,t} * 0,5 \left({}_nW_x^t + {}_nW_{x+n}^t \frac{{}_nL_x^{t-5,t}}{{}_nL_{x+n}^t} \right) \quad (6)$$

Ao rearranjar a equação (6), pode-se colocar cada grupo etário de mulheres em evidência. Assim, é possível criar uma função de nascimentos composta pelas taxas específicas médias de fecundidade do quinquênio.

$$B^{t-5,t} = 5 * 0,5 \sum_{15}^{54} \left[({}_nF_x^{t-5,t} * {}_nW_x^t) + \left({}_nF_{x+n}^{t-5,t} * {}_nW_{x+n}^t \frac{{}_nL_x^{t-5,t}}{{}_nL_{x+n}^t} \right) \right] \quad (7)$$

$$B^{t-5,t} = 5 * 0,5 \left[\begin{aligned} & \left({}_5F_{15}^{t-5,t} * {}_5W_{15}^t \right) + \left({}_5F_{15}^{t-5,t} * {}_5W_{20}^t \frac{{}_5L_{15}^{t-5,t}}{{}_5L_{20}^{t-5,t}} \right) + \\ & \left({}_5F_{20}^{t-5,t} * {}_5W_{20}^t \right) + \left({}_5F_{20}^{t-5,t} * {}_5W_{25}^t \frac{{}_5L_{20}^{t-5,t}}{{}_5L_{25}^{t-5,t}} \right) + \\ & \left({}_5F_{25}^{t-5,t} * {}_5W_{25}^t \right) + \left({}_5F_{25}^{t-5,t} * {}_5W_{30}^t \frac{{}_5L_{25}^{t-5,t}}{{}_5L_{30}^{t-5,t}} \right) + \\ & \left({}_5F_{30}^{t-5,t} * {}_5W_{30}^t \right) + \left({}_5F_{30}^{t-5,t} * {}_5W_{35}^t \frac{{}_5L_{30}^{t-5,t}}{{}_5L_{35}^{t-5,t}} \right) + \\ & \left({}_5F_{35}^{t-5,t} * {}_5W_{35}^t \right) + \left({}_5F_{35}^{t-5,t} * {}_5W_{40}^t \frac{{}_5L_{35}^{t-5,t}}{{}_5L_{40}^{t-5,t}} \right) + \\ & \left({}_5F_{40}^{t-5,t} * {}_5W_{40}^t \right) + \left({}_5F_{40}^{t-5,t} * {}_5W_{45}^t \frac{{}_5L_{40}^{t-5,t}}{{}_5L_{45}^{t-5,t}} \right) + \\ & \left({}_5F_{45}^{t-5,t} * {}_5W_{45}^t \right) + \left({}_5F_{45}^{t-5,t} * {}_5W_{50}^t \frac{{}_5L_{45}^{t-5,t}}{{}_5L_{50}^{t-5,t}} \right) \end{aligned} \right] \quad (8)$$

⁴ Formulação construída para uma população fechada.

$$B^{t-5,t} = 5 * 0,5 \begin{bmatrix} 5W_{15}^t * 5F_{15}^{t-5,t} + \\ 5W_{20}^t * \left(5F_{20}^{t-5,t} * 5F_{15}^{t-5,t} \frac{5L_{15}^{t-5,t}}{5L_{20}^{t-5,t}} \right) + \\ 5W_{25}^t * \left(5F_{25}^{t-5,t} * 5F_{20}^{t-5,t} \frac{5L_{20}^{t-5,t}}{5L_{25}^{t-5,t}} \right) + \\ 5W_{30}^t * \left(5F_{30}^{t-5,t} * 5F_{25}^{t-5,t} \frac{5L_{25}^{t-5,t}}{5L_{30}^{t-5,t}} \right) + \\ 5W_{35}^t * \left(5F_{35}^{t-5,t} * 5F_{30}^{t-5,t} \frac{5L_{30}^{t-5,t}}{5L_{35}^{t-5,t}} \right) + \\ 5W_{40}^t * \left(5F_{40}^{t-5,t} * 5F_{35}^{t-5,t} \frac{5L_{35}^{t-5,t}}{5L_{40}^{t-5,t}} \right) + \\ 5W_{45}^t * \left(5F_{45}^{t-5,t} * 5F_{40}^{t-5,t} \frac{5L_{40}^{t-5,t}}{5L_{45}^{t-5,t}} \right) + \\ 5W_{50}^t * \left(5F_{45}^{t-5,t} \frac{5L_{45}^{t-5,t}}{5L_{50}^{t-5,t}} \right) \end{bmatrix} \quad (9)$$

O mesmo conjunto de equações pode ser reduzido a apenas uma equação geral, (10), e aplicada uma função de sobrevivência, o que estima o número de pessoas entre 0 e 4 anos no tempo t (11).

$$B^{t-5,t} = 5 \sum_{i=15}^{54} 5W_x^t \left[0,5 \left(5F_x^{t-5,t} + 5F_{x-5}^{t-5,t} * \frac{5L_{x-5}^{t-5,t}}{5L_x^{t-5,t}} \right) \right] \quad (10)$$

$$5N_0^t = \sum_{i=15}^{54} 5W_x^t \left[0,5 \left(5F_x^{t-5,t} + 5F_{x-5}^{t-5,t} * \frac{5L_{x-5}^{t-5,t}}{5L_x^{t-5,t}} \right) \right] * \frac{5L_0^{t-5,t}}{1_0^{t-5,t}} \quad (11)$$

Relação entre taxa de fecundidade total e razão criança-mulher

A leitura da equação (11) permite estimar o número de crianças entre 0 e 4 anos de idade, no ano t (em 2000) como o número de nascimentos gerados pelas mulheres em idade reprodutiva no último quinquênio e que sobreviveram a esse intervalo. A contribuição de cada mulher viva no tempo t para esses nascimentos corresponde à média entre o risco corrente do grupo quinquenal ao qual pertence e do grupo etário imediatamente anterior, mesmo aquelas que não sobreviveram ao final do quinquênio.

Em outros termos, as mulheres que estão vivas no grupo quinquenal $x, x+n$, no tempo t , contribuirão metade do tempo entre $t-5$ e t com a fecundidade específica de seu grupo etário, ao passo que, na outra metade do tempo, a contribuição foi feita na ordem da fecundidade específica do grupo etário imediatamente inferior. Entretanto, entre $t-5$ e t , algumas mulheres faleceram e, para reconstituir os nascimentos oriundos dessas mulheres, retroprojeta-se a contribuição das mesmas para o número de nascimentos produzidos entre $t-5$ e t ao se multiplicar a taxa específica do grupo etário inferior pelo inverso da razão de sobrevivência.

Por fim, para definir o número de nascimentos sobreviventes, multiplica-se o número médio estimado de nascimentos entre $t-5$ e t pelo número de pessoas-ano extraído da tabela de vida do período $t-5$ e t (${}_5L_0^{t-5,5}$). A partir da manipulação algébrica da equação (11) e considerando ${}_5\varphi_x = 5F_x / TFT$ (onde TFT é a taxa de fecundidade total), pode-se reescrever a equação da seguinte forma:

$${}_5N_0^t = \frac{TFT^{t-5,t}}{5} * \frac{{}_5L_0^{t-5,t}}{l_0^{t-5,t}} \sum_{i=15}^{49} {}_5W_x^t \left[0,5 \left({}_5\varphi_x^{t-5,t} + {}_5\varphi_{x-5}^{t-5,t} * \frac{{}_5L_x^{t-5,t}}{{}_5L_{x-5}^{t-5,t}} \right) \right] \quad (12)$$

Se cada lado da igualdade for dividido pelo total de mulheres sobreviventes em t e que contribuíram para a geração de nascimentos (W), tem-se:

$$\frac{{}_5N_0^t}{{}_{35}W_{15}^t} = TFT^{t-5,t} * \frac{{}_5L_0^{t-5,t}}{{}_5l_0^{t-5,t}} \sum_{i=15}^{49} \frac{{}_5W_x^t}{{}_{35}W_{15}^t} \left[0,5 \left({}_5\varphi_x^{t-5,t} + {}_5\varphi_{x-5}^{t-5,t} * \frac{{}_5L_x^{t-5,t}}{{}_5L_{x-5}^{t-5,t}} \right) \right] \quad (13)$$

Onde $\frac{{}_5N_0^t}{{}_{35}W_{15}^t}$ é uma medida conhecida como razão criança-mulher.⁵ Essa razão é utilizada em diversos trabalhos, sendo uma ferramenta poderosa para estimar relações demográficas diversas, relacionando ora os nascimentos, ora o número de nascimentos sobreviventes, com a população feminina em idade reprodutiva (BRASS, 1953, 1954; DUCHESNE, 1989). Se chamarmos o termo $\sum_{i=15}^{49} \frac{{}_5W_x^t}{{}_{35}W_{15}^t} \left[0,5 \left({}_5\varphi_x^{t-5,t} + {}_5\varphi_{x-5}^{t-5,t} * \frac{{}_5L_x^{t-5,t}}{{}_5L_{x-5}^{t-5,t}} \right) \right]$ de λ e $\frac{{}_5L_0^{t-5,t}}{{}_5l_0^{t-5,t}}$ de $S_0^{t-5,t}$, tem-se:

$$\frac{{}_5N_0^t}{{}_{35}W_{15}^t} = TFT^{t-5,t} * S_0^{t-5,t} * \lambda \quad (14)$$

$$TFT^{t-5,t} = \frac{{}_5N_0^t}{{}_{35}W_{15}^t} * \frac{1}{S_0^{t-5,t}} * \frac{1}{\lambda} \quad (15)$$

Tomando o logaritmo natural de ambos os lados da equação (15):

$$\ln TFT^{t-5,t} = \ln \frac{{}_5N_0^t}{{}_{35}W_{15}^t} + \ln \frac{1}{S_0^{t-5,t}} + \ln \frac{1}{\lambda} \quad (16)$$

Se reescrevermos $\ln \frac{1}{S_0^{t-5,t}} + \ln \frac{1}{\lambda}$ como c e considerarmos uma constante θ com o valor de uma unidade, reescreve-se a equação (16) da seguinte forma:

$$\ln TFT^{t-5,t} = \theta \ln \frac{{}_5N_0^t}{{}_{35}W_{15}^t} + c \quad (17)$$

A equação 17 descreve uma função linear com o coeficiente de inclinação igual a um. Entretanto, o coeficiente θ pode apresentar valores diferentes da unidade em situações de subenumeração ou sobreenumeração censitária, tanto no numerador quanto no denominador. Ou seja:

$$\theta \ln \frac{{}_5N_0^t}{{}_{35}W_{15}^t} = \ln \frac{{}_5N_0^t}{{}_{35}W_{15}^t} \quad \forall \theta \neq 1 \quad (18)$$

Onde ${}_5N_0^t$ e ${}_{35}W_{15}^t$ são, respectivamente, o número de crianças entre 0 e 4 anos e o de mulheres de 15 a 50 anos enumeradas no censo.

Se considerarmos, como pressuposto, que as mulheres em período reprodutivo possuem erro de enumeração quase nulo e que a população é fechada, tem-se que o erro de enumeração incide apenas na população de 0 a 4 anos de idade. Assim sendo, a equação (18) descreve um modelo linear que pode ser estimado por mínimos quadrados. É possível obter os valores de TFT e $\frac{{}_5N_0^t}{{}_{35}W_{15}^t}$ para recortes territoriais menores, segundo a estimação do modelo geral, sob a condição de que os coeficientes médios convirjam para o valor real dos parâmetros da grande área. Nesse sentido, pressupõe-se que a dimensão das pequenas áreas não interfere de forma substantiva na estimação dos parâmetros.

⁵ Para essa formulação ver também Schmetmann e Hauer (2017).

Se a condição na equação (18) é verdadeira, então o número de pessoas entre 0 e 4 anos no tempo t é dado por:

$${}_5N_0^t = {}_{35}W_{15}^t e^{\theta \ln \frac{{}_5N_0^t}{{}_{35}W_{15}^t}} \quad (19)$$

Nesse caso, o valor de θ contribui para a correção da razão criança-mulher. Para estimar o número total de crianças de 0 a 4 anos, pode-se partir do pressuposto de que a população adulta feminina é bem enumerada, ou mesmo corrigir a população escolhendo a melhor enumeração entre os dois censos para qualificar a estimativa.⁶ Contudo, se sabemos que:

$${}_5N_0^t = \frac{{}_5N_0^t}{{}_5N_0^t} {}_5N_0^t = \alpha {}_5N_0^t \quad (20)$$

Onde α é o fator de correção linear da enumeração censitária em t . Logo, tem-se que:

$$\alpha = \left(\frac{{}_5N_0^t}{{}_{35}W_{15}^t} \right)^{-1} e^{\theta \ln \frac{{}_5N_0^t}{{}_{35}W_{15}^t}} \quad (21)$$

Onde $1/\alpha$ é o grau de enumeração da população entre 0 e 4 anos de idade no período t . Neste trabalho, foram estimadas as razões criança-mulher em 2000 e as funções de fecundidade para o quinquênio 1995-2000, a fim de se estimar o parâmetro θ , segundo a equação (18), por meio de mínimos quadrados ordinários. Foram utilizadas como unidades de análise os estados brasileiros e as mesorregiões.⁷

Método de Preston para estimação da taxa bruta de natalidade por meio de duas distribuições etárias

Preston (1983) propõe um método para identificar um fator de correção para o nível de mortalidade de uma função padrão predefinida, bem como gerar uma estimativa sobre a taxa bruta de natalidade. Para tanto, o autor lança mão de relações de populações não estáveis, tal como descrito a seguir. Sendo a estrutura etária uma resultante da função da taxa bruta de natalidade, de uma função de sobrevivência e da diferença de tamanho entre as coortes, dada pela variável r (extraída da taxa de crescimento entre dois censos, segundo os grupos etários), então, têm-se que:

$$c_{(a,t)} = b_{(t)} e^{-\int_0^a r(x,t) dx} p_{(a)} \quad (22)$$

$$\frac{1}{p_{(a)}} = \frac{b_{(t)} e^{-\int_0^a r(x,t) dx}}{c_{(a,t)}} \quad (23)$$

Em seguida, Preston (1983) manipula o modelo relacional de Brass, isolando a função de sobrevivência.

⁶ Isso pode ser feito utilizando a razão entre os fatores de enumeração censitária fornecidos pela técnica *General Growth Balance* (HILL, 1987; UNITED NATIONS, 2002).

⁷ O nível de fecundidade do Brasil, estados e mesorregiões foi estimado a partir da definição de uma curva logística que passa pelos níveis de fecundidade de 2000 e 2010 corrigidos pela técnica P/F de Brass. O nível mais alto foi definido com sete filhos por mulher e o mais baixo 1,1. Já o padrão foi estimado a partir da extrapolação entre as funções estimadas para 2000 e 2010, utilizando o modelo relacional de Gompertz tradicional.

$$\ln \frac{q(a)}{p(a)} = \alpha + B \ln \frac{q_s(a)}{p_s(a)} \quad (24)$$

$$\frac{1-p(a)}{p(a)} = e^\alpha \left(\frac{q_s(a)}{p_s(a)} \right)^B \quad (25)$$

$$\frac{1}{p(a)} = e^\alpha \left(\frac{q_s(a)}{p_s(a)} \right)^B + 1 \quad (26)$$

Se e^α pode ser descrito como K (indicador de diferença entre os níveis de mortalidade) e assume-se que B possui o valor 1, ou seja, aceita-se que a função padrão possui a mesma estrutura da função da população para a qual se deseja estimar o nível, tem-se que:

$$\frac{1}{p(a)} = 1 + K \frac{q_s(a)}{p_s(a)} \quad (27)$$

Igualando as equações (26) e (27), uma vez que as duas definem $\frac{1}{p(a)}$, determina-se uma função linear com dois parâmetros, K/b (inclinação) e $1/b$ (intercepto).

$$\frac{b_{(t)} e^{-\int_0^a r(x,t) dx}}{c_{(a,t)}} = 1 + K \frac{q_s(a)}{p_s(a)} \quad (28)$$

$$\frac{e^{-\int_0^a r(x,t) dx}}{c_{(a,t)}} = \frac{1}{b_{(t)}} + \frac{K}{b_{(t)}} \frac{q_s(a)}{p_s(a)} \quad (29)$$

Logo, a equação (29) é uma função linear com dois parâmetros. Assim, a partir do produto da razão de sobrevivência para o primeiro grupo da população total, do intervalo n (entre $t-5$ e t) e da taxa bruta de natalidade, tem-se a população estimada para o primeiro grupo.

$${}_5N_0^t = {}_\infty N_0^t \frac{{}_5L_0^t}{{}_5l_0^t} n b_{(t)} \quad (30)$$

A estimativa da variável r foi feita a partir da relação entre os Censos Demográficos de 1991 e 2000. Segundo as taxas de crescimento por grupo etário, foi estimada a população de 1995 para estimar a população em cinco anos. Assim, a taxa bruta de natalidade foi estimada não para o tempo t , mas sim para o quinquênio 1995-2000. Para detalhes da operacionalização, ver Preston (1983).

Resultados

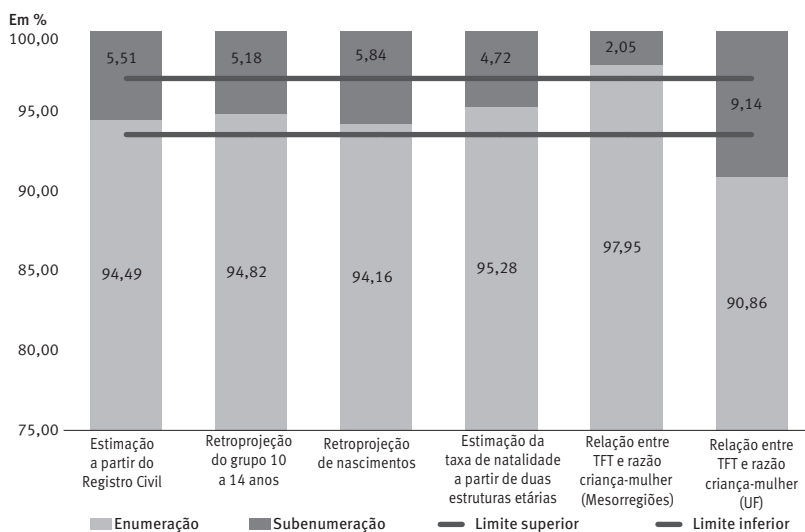
Foram utilizadas cada uma das cinco técnicas na estimação da população entre 0 e 4 anos, de 2000, para o total do Brasil. Dentre as técnicas, a relação entre a taxa de fecundidade total e razão criança-mulher foi aplicada para as unidades da federação e mesorregiões, como já apontado. Essa técnica, bem como aquela de Preston (1983), possui a vantagem de oferecer um intervalo de confiança da estimativa.

Todavia, o intervalo de confiança da relação entre a taxa de fecundidade total e a razão criança-mulher apresentou valores pouco satisfatórios e inconsistentes entre si.

O intervalo (95% de confiança) da enumeração censitária, ao utilizar as mesorregiões, ficou entre 97,87% e 105,77%, sendo inconclusivo e levantando a possibilidade de haver sobreenumeração da população de 0 a 4 anos em 2000, algo pouco factível. Já para as unidades da federação, esse intervalo, com o mesmo nível de confiança, ficou entre 83,10% e 99,35%, muito dilatado e inconclusivo.⁸

O Gráfico 1 mostra a síntese dos resultados, com o intervalo de confiança do método proposto por Preston (1983). A construção do modelo linear entre o nível de fecundidade e a razão-criança mulher apresenta resultados muito diversos, para além daqueles discutidos sobre o intervalo de confiança. A estimação a partir do uso do Registro Civil e a retroprojeção do grupo de 10 a 14 anos, utilizadas pelo IBGE, mostram resultados razoáveis e muito próximos.

GRÁFICO 1
Estimativas da enumeração censitária entre 0 e 4 anos, segundo técnicas de estimação
Brasil – 2000



Fonte: IBGE, Censos Demográficos de 1991, 2000 e 2010; Registro Civil, 1995-2000. Elaboração dos autores a partir da aplicação dos métodos supracitados.

A retroprojeção de nascimentos, por seu turno, converge para valores similares aos das técnicas anteriormente citadas. Todos eles, no entanto, localizam-se dentro do intervalo de confiança do método de estimação da taxa bruta de natalidade, por meio das estruturas etárias, mesmo que a média estimada seja de 4,72%, abaixo das demais. Esse resultado é efeito, em parte, da não correção do nível de mortalidade pelo fator K do próprio método, que reduziria o nível, aumentando a sobrevivência.

⁸ Todos os coeficientes da regressão foram significativos a 0,001, com R2 igual a 0,9512, no modelo com as unidades da federação, e 0,9343, para aquele com as mesorregiões.

Discussão

Em linhas gerais, podem ser destacadas algumas questões sobre cada uma das técnicas utilizadas, o que abre possibilidade de novas tentativas e ajustes com os mesmos métodos, a ser feito em trabalhos posteriores. A retroprojeção do grupo etário de 10 a 14 anos e a utilização do Registro Civil dos cinco anos que antecedem o ano do censo em questão oferecem estimativas confiáveis e consistentes entre si, porém mantêm as limitações de sub-registro e do registro tardio – no caso do Registro Civil – e subenumeração – no caso do segundo censo.

Com relação ao registro tardio, apesar da sua existência, destaca-se que ele tende a diminuir com o passar dos anos. Ao se adotar os sete anos posteriores ao ano de nascimento da criança, estar-se-ia cometendo um baixo erro, por falta, no total de nascimentos, já que a certidão de nascimento da criança é uma pré-condição para que ela seja matriculada na escola. Mesmo assim, pode ser pouco operacional aguardar sete anos após a execução do censo demográfico para a correção da enumeração, o que torna a técnica pouco atrativa no curto prazo.⁹

Apesar de também estimar um intervalo de confiança, a relação entre razão criança-mulher e taxa de fecundidade total parece sofrer efeito de composição das unidades de análise, objeto da estimação. Diferenças dos níveis de mortalidade na infância, bem como a agregação diferente entre as unidades de análise, suavizariam grande parte da variância interna às mesmas. Além disso, unidades menores não são necessariamente independentes quanto às variáveis em tela, o que gera distorções no modelo.

As estimativas baseadas na retroprojeção de nascimentos e os demais resultados acima mencionados convergem, com resultados ligeiramente superiores quanto à subenumeração. A limitação, contudo, permanece com a estimação da função de fecundidade dentro dos cinco anos que antecedem o censo, fortemente dependente dos níveis estimados a partir, pelo menos, de dois censos, bem como a enumeração feminina adulta.

Esses vieses podem ter direções opostas, caso o nível de fecundidade seja sobrestimado e haja subenumeração feminina. Além disso, o padrão de fecundidade é crucial, pois afeta diretamente o volume de nascimentos por meio de sua ponderação pela distribuição etária. Assim, a idade média da função de fecundidade torna-se um importante componente de projeção dos nascimentos.

Por fim, o método desenvolvido por Preston (1983) ofereceu um resultado consistente e possui a vantagem de definir um intervalo de confiança para a estimativa. Entretanto, o método é fortemente influenciado pela enumeração censitária e pela distribuição etária dos censos analisados. Testes com censos diferentes e tabelas de vida de outras fontes podem ajudar na simulação de erros e identificação de vantagens e limitações. No presente

⁹ Outra possibilidade é assumir o pressuposto de que o percentual de registros por ano após o nascimento é constante. Assim sendo, assume-se que a distribuição relativa do registro de nascimento entre um ano t , ao longo do período t a $t+7$ seria igual ao período $t+7$ a $t+14$.

exercício, o método ofereceu resultados satisfatórios, inclusive para a correção do nível de mortalidade utilizado, neste caso as tabelas estimadas pelo IBGE para o período.

Referências

BRASS, W. The derivation of fertility and reproduction rates from restricted data on reproductive histories. *Population Studies*, v. 7, n. 2, p. 137-166, 1953.

_____. The estimation of fertility rates from ratios of total to first births. *Population Studies*, v. 8, n. 1, p. 74-87, 1954.

DUCHESNE, L. Proyecciones de población, por sexo y edad, para áreas intermedias y menores: método relación de cohortes. In: GRANADOS, M. P. **Métodos para proyecciones subnacionales de población**. Bogotá: Centro Latinoamericano de Demografía, 1989. p. 71-126.

HILL, K. Estimating census and death registration completeness. *Asian and Pacific Census Forum*, v. 1, n. 3, p. 8-13, 23-24, 1987.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Censo demográfico 2000: resultados da pesquisa de avaliação da cobertura da coleta**. Rio de Janeiro: IBGE, Coordenação Técnica do Censo Demográfico, 2003 (Textos para discussão, n. 9).

_____. **Projeções da população: Brasil e unidades da federação**. Rio de Janeiro: IBGE, 2013 (Série Relatórios Metodológicos, 40).

_____. **Projeções da população: Brasil e unidades da federação**. 2. ed. Rio de Janeiro: IBGE, 2018 (Série Relatórios Metodológicos, 40).

OLIVEIRA, L. C. de S.; FREITAS, M. P. S. de; DIAS, M. R. M. L.; NASCIMENTO; C. M. F.; MATTOS, E. da S.; RAMALHO JÚNIOR, J. J. A. **Censo demográfico 2000: resultados da pesquisa de avaliação da cobertura da coleta**. Rio de Janeiro: IBGE, Coordenação Técnica do Censo Demográfico, 2003.

PRESTON, S. H. An integrated system for demographic estimation from two age distributions. *Demography*, v. 20, n. 2, p. 213-226, 1983.

SCHMERTMANN, C.; HAUER, M. Bayesian estimation of total fertility from a population's age-sex Distribution. 2017. Disponível em: <<https://osf.io/preprints/socarxiv>>.

UNITED NATIONS. **Manual III. Methods for population projections by sex and age**. United Nations publication, Sales n. 56.XIII.3, 1956.

_____. **Methods for estimating adult mortality**. New York: UN, 2002 (Working Paper, n. ESA/P/WP.175).

Sobre os autores

Reinaldo Onofre Santos é doutorando e mestre em Demografia no Centro de Desenvolvimento e Planejamento Regional (Cedeplar) da Universidade Federal de Minas Gerais (UFMG).

Guilherme Quaresma Gonçalves é doutorando em Demografia no Centro de Desenvolvimento e Planejamento Regional (Cedeplar) da Universidade Federal de Minas Gerais (UFMG) e mestre em Ciência Política pelo Departamento de Ciência Política (DCP) da UFMG.

Endereço para correspondência

Reinaldo Onofre Santos

Cedeplar – Centro de Desenvolvimento e Planejamento Regional da UFMG
Faculdade de Ciências Econômicas – 3º andar
Avenida Presidente Antônio Carlos, 6627, Pampulha
31270-901 – Belo Horizonte-MG, Brasil

Guilherme Quaresma Gonçalves

Cedeplar – Centro de Desenvolvimento e Planejamento Regional da UFMG
Faculdade de Ciências Econômicas – 3º andar
Avenida Presidente Antônio Carlos, 6627, Pampulha
31270-901 – Belo Horizonte-MG, Brasil

Abstract

Methods for estimation of subenumeration of the 0-4 year old in the 2000 Brazilian Demographic Census

Our aim is to apply techniques which contribute to identify and correct the census under-enumeration for people aged 0-4 years old, using population enumerated in the 2000 Brazilian census as reference for estimations. Usually, this group has lower enumeration than others, such as adult population. Five techniques were applied to the 0-4 year old group estimation and results indicate that, although usual techniques have offered satisfactory results, alternative forms have great potential to be explored.

Keywords: Demographic census. Under-enumeration. Demographic analysis techniques.

Resumen

Métodos para la estimación de la subenumeración del grupo cero a cuatro años en el Censo Demográfico Brasileño del año 2000

El objetivo de este trabajo es la aplicación de técnicas que contribuyen en la identificación y corrección de la subenumeración censal del grupo etario de cero a cuatro años de edad, sobre la base de la población enumerada en el Censo Demográfico de 2000. Frecuentemente, este grupo posee baja enumeración, si se lo compara con la de otros grupos etarios de la población adulta. Para cumplir el objetivo, fueron aplicadas cinco técnicas para la estimación del grupo de cero a cuatro años, y los resultados señalan que, pese a que las técnicas comúnmente usadas ofrecen resultados satisfactorios, algunas formas alternativas también muestran gran potencial a ser explorado.

Palabras clave: Censos demográficos. Subenumeración. Técnicas de análisis demográfico.

Recebido para publicação em 08/09/2018

Aceito para publicação em 13/11/2018