



Revista Latinoamericana de Población

ISSN: 2393-6401

revista.relap@gmail.com

Asociación Latinoamericana de Población

Uruguay

Miranda de Araújo Freire, Flávio Henrique; Gonzaga,
Marcos Roberto; Forte Gomes, Marília Miranda
Projeções populacionais por sexo e idade para pequenas áreas no Brasil
Revista Latinoamericana de Población, vol. 13,
núm. 26, 2020, Novembro-Dezembro, pp. 124-149
Asociación Latinoamericana de Población
Uruguay

DOI: <https://doi.org/10.31406/relap2020.v14.i1.n26.6>

Disponível em: <https://www.redalyc.org/articulo.oa?id=323861646007>

- Como citar este artigo
- Número completo
- Mais informações do artigo
- Site da revista em redalyc.org

UABM
redalyc.org

Sistema de Informação Científica Redalyc

Rede de Revistas Científicas da América Latina e do Caribe, Espanha e Portugal

Sem fins lucrativos acadêmica projeto, desenvolvido no âmbito da iniciativa
acesso aberto

Projeções populacionais por sexo e idade para pequenas áreas no Brasil

Sex and Age-Specific Population Projections for Small Areas in Brazil

Flávio Henrique Miranda de Araújo Freire

Orcid: 0000-0002-7416-9947

fhfreire@ccet.ufrn.br

Universidade Federal do Rio Grande do Norte, Brasil

Marcos Roberto Gonzaga

Orcid: 0000-0002-6088-3453

marcosrg@ccet.ufrn.br

Universidade Federal do Rio Grande do Norte, Brasil

Marília Miranda Forte Gomes

Orcid: 0000-0001-8584-9676

mariliamfg@gmail.com

Universidade de Brasília, Brasil

Resumo

O método de relação de coortes é um método demográfico utilizado para projeções populacionais por sexo e idade para pequenas áreas. Porém, devido a instabilidade dos eventos demográficos em áreas menores, o método produz resultados pouco razoáveis para alguns grupos de idade. Neste artigo aplicamos um estimador contração, bayesiano empírico, para reduzir a instabilidade nas estimativas dos fatores diferenciais de crescimento do método em áreas menores. Ao aplicarmos o método de relação de coortes, ajustado pelo bayesiano empírico, para projeções populacionais municipais no Brasil entre 2000 e 2010, observamos uma redução na variabilidade das estimativas por idade. Além disso, o método reduziu o erro percentual absoluto médio das populações projetadas nesse decênio comparativamente aos resultados obtidos com a aplicação do método original proposto por Duchesne (1989), especialmente nos grupos etários de adultos e idosos.

Palabras-chave

Projeção
Demografia
Pequenas áreas
Bayesiano
empírico
Método da relação de Coortes

Abstract

The cohort ratio method is a demographic approach used for sex and age-specific population projections in small areas. However, due to the instability of the demographic events in small areas, this method

produces unreasonable results in some age groups. In this paper, we apply a Bayesian spatial statistic technique (empirical Bayes) to reduce instability when estimating the cohort changes' index of the cohort ratio method for small areas. When applying the method for population projections in some Brazilian small areas, from 2000 to 2010, we observe a reduction on the variability of the estimations. In addition, our method reduced the error of the projections from 2000 to 2010 when compared to the method originally proposed by Duchesne (1989), especially on the last age groups.

Keywords

Projection
Demography
Small areas
Empirical Bayes
Cohort ratio
method

Recibido: 18/01/2019
Aceptado: 05/08/2019

Introdução

Desde a segunda metade do século passado o Brasil vem experimentando importantes mudanças demográficas, que ocorreram em ritmos diferentes ao longo do território, e que teve variações ao longo do tempo. Todas as regiões do país experimentaram significantes declínios nas taxas de mortalidade, especialmente nas primeiras idades, e diferentes transformações nas taxas específicas de mortalidade por causas de morte (Araújo, 2012; Castro e Simões, 2009; Prata, 1992; Schramm et al., 2004). As taxas de fecundidade reduziram em todas as regiões em diferentes ritmos de declínio (Potter, Schmertmann, Assunção e Cavenagui, 2010; Wong e Carvalho, 2006). Os fluxos migratórios do Nordeste para o Sudeste do país se tornaram menos intensos ao passo que se intensificaram os fluxos dentre uma mesma unidade federativa (UF) ou região (Cunha e Baeninger, 2005; Moura, 1980; Oliveira e Oliveira, 2011; Pacheco e Patarra, 1997). Todas essas mudanças modificaram o perfil da população das unidades federativas, e, *são ainda mais impactantes*, quando as áreas geográficas de análise diminuem em tamanho, tornando-se mais complexa a estimação e a projeção de taxas demográficas.

Estimativas e projeções populacionais, por variáveis demográficas (ex. sexo e idade) ou níveis geográficos mais desagregados, são importantes para o delineamento e planejamento de políticas e alocação de recursos públicos em diversas áreas (Barros e Cavenaghi, 2017; Rees, 1989; Santos e Barbieri, 2015). Porém, estimativas precisas de populações subnacionais por sexo e idade, especialmente em países em desenvolvimento, ainda são desafios para demógrafos e estudiosos de população, especialmente em áreas com limitações na qualidade dos registros vitais e contagem de população (Justino, Freire e Lucio, 2012; Paes e Albuquerque, 1999). Para lidar com essa demanda, é necessário o desenvolvimento ou aprimoramento de métodos e técnicas que permitem incorporar, nas estimativas de população, possíveis heterogeneidades nos eventos demográficos em áreas menores.

Os métodos tradicionalmente utilizados para estimar e projetar população em pequenas áreas estão centrados nas tendências históricas do crescimento populacional, na relação que algumas variáveis socioeconômicas poderiam ter com o número de habitantes de determinada área ou, quando possível, nas três componentes que regem toda a dinâmica demográfica local (Brito, Cavenagui e Jannuzzi, 2010; Granados, 1989; Jannuzzi, 2006; Jardim, 2000; Rao, 2003; Smith, Tayman e Swanson, 2001; Sprague, 2013; Waldvogel, 1998). Porém, esses métodos tradicionais não incorporam um aspecto importante em projeções populacionais de pequenas áreas: a configuração e correlação espacial da área em estudo.

O *método de relação de coortes* é tradicionalmente empregado para projeções populacionais por sexo e grupos de idade em pequenas áreas no Brasil (Assunção, 2002; Ipardes e IBGE, 1999). O método calcula um fator diferencial de crescimento por coortes etárias das populações de cada área menor e, em seguida, extrapola o crescimento das coortes de cada área menor em relação ao crescimento das mesmas coortes etárias numa área maior que compreende, geograficamente, todas as áreas menores (Duchesne, 1989). A principal vantagem desse método é a incorporação das mudanças no tamanho da população por coortes de idade de cada área menor em relação às mudanças nas coortes equivalentes da área maior. Ao utilizar a razão entre as relações de sobrevivência das coortes etárias das áreas menores e, as relações de sobrevivência das coortes equivalentes na área maior, denominada fator diferencial de crescimento da coorte,¹ assume-se que o método incorpora parte da lógica que rege a dinâmica demográfica da região. Dentre outras vantagens, o método projeta não só o volume total da população de cada área menor, mas também uma estrutura etária populacional de cada área menor condizente com a estrutura etária da área maior. Por essas e outras vantagens o método de relação das coortes pode ser considerado como uma variante do método de coortes componentes necessitando, porém, de um menor refinamento de dados para ser implementado, conforme descrevemos mais adiante.

Por outro lado, uma das principais limitações do método se refere ao fato de os fatores diferenciais de crescimento por coortes etárias serem tomados como constante por todo o período da projeção. Esses fatores são calculados diretamente das informações censitárias entre dois censos, mormente com observações dos dois últimos censos disponíveis. Além disso, não é raro observar a ocorrência de valores extremos (discrepantes) nos fatores diferenciais de crescimento em determinados grupos de idade e em algumas pequenas áreas, produzindo estimativas pouco razoáveis em termos de estrutura etária e do volume da população (Alves, 2016). Manter esses valores discrepantes nos fatores diferenciais de crescimento por todo o período da projeção, geraria erros acumulativos a cada quinquênio, resultando em projeções pouco razoáveis quanto menor for a área em questão.

Dentre outras limitações menores, ressalta-se o fato de não haver uma perfeita igualdade entre a soma das populações por idade entre as áreas menores em relação às mesmas idades na área maior. No que diz respeito a outras limitações menores, sugerimos a leitura de Duchesne (1989).

Entretanto, em detrimento das limitações do método de relação de coortes, seja pela qualidade dos dados ou pela dinâmica dos eventos demográficos em pequenas áreas, trata-se de um método demográfico que incorpora, ainda que indiretamente, algumas nuances das componentes demográficas que não podem ser previstas ou projetadas por puras equações matemáticas ou modelos estatísticos.

Há uma vasta variedade de métodos de projeções populacionais de pequenas áreas, alguns baseados em relações que tentam levar em consideração a dinâmica demográfica, ainda que indiretamente, outros que tentam relacionar o crescimento populacional com variáveis correlacionadas ao incremento ou decréscimo da população (variáveis sintomáticas) e os métodos por extrapolação matemática

1 Para efeito de simplificação, muitas vezes esses fatores diferenciais de crescimento das coortes serão chamados de: fator (ou fatores) K.

(Cavenaghi, 2012; Granados, 1989; Jannuzzi, 2006; Rao, 2003; Santos & Barbieri, 2015; Smith et al., 2001; Sprague, 2013; Waldvogel, 1998).

Em alguns países latino-americanos, como Argentina, Chile e Peru, utilizam-se um conjunto de procedimentos que envolve a projeção da participação relativa da população da área menor com relação à área maior usando uma função logística. A população da área maior deve ser projetada anteriormente, assim como ocorre no método relações de coortes. A partir dessa população total projetada para a área menor, a estrutura etária é obtida a partir de um processo iterativo com tabelas de contingência para ajustar a distribuição da população por idade das áreas menores à distribuição etária da área maior, em determinado período da projeção. No manual VIII das Nações Unidas (Naciones Unidas, 1975) há uma descrição do procedimento. Arriaga (2001) também traz detalhes sobre esta metodologia.

No Brasil, Corrêa e colaboradores (2011) utilizaram uma função logística para projetar a participação relativa da população de áreas menores com relação à uma área maior, usando o arcabouço metodológico do método *apportionment method* ou AiBi como ficou conhecido no Brasil (Madeira e Simões, 1972).

Neste artigo, privilegiamos o método relações de coortes, por entender que sua distribuição etária é obtida a partir de um mecanismo, com as relações de sobrevivência, que considera a dinâmica demográfica de forma mais direta a partir das componentes que regem essa dinâmica populacional. Além disso, combinamos técnicas de estatística espacial com aplicação do método de relação das coortes com o objetivo de amenizar uma das principais limitações do método proposto por Duchesne (1989). Propomos a utilização do *modelo bayesiano empírico*, que incorpora a distribuição espacial das áreas vizinhas ou homogêneas *àquela área menor em questão* e reduz as flutuações excessivas no cálculo dos fatores diferenciais de crescimento do método de relação de coortes (Duchesne, 1989).

A solução apresentada por Duchesne (1989) para reduzir a alta variabilidade nos fatores diferenciais de crescimento por grupos de idade seria algum tipo de ajuste ou suavização matemática nesses fatores ao longo dos grupos etários. Consideramos que a incorporação do estimador contração do tipo bayesiano empírico produz uma suavização nos fatores diferenciais de crescimento por coortes etárias, diminuindo sua variabilidade e gerando melhores resultados na projeção populacional pelo método relações de coortes.

Assim, neste artigo, além de apresentar os algoritmos para aplicação do método de relação de coortes e o desenvolvimento do estimador bayesiano empírico para o cálculo do fator diferencial de crescimento, apresentamos uma análise dos erros de projeção nas seguintes situações: i) aplicação do método de relação de coortes originalmente proposto por Duchesne (1989), e ii) aplicação do método de relação de coortes com ajuste pelo estimador bayesiano empírico no cálculo dos fatores diferenciais de crescimento das coortes. Finalmente, apresentamos os resultados da projeção dos 5.570 municípios do Brasil entre 2010 e 2030. Desenvolveu-se um algoritmo em R para implementação do método de relação das coortes com ajuste pelo estimador bayesiano empírico, que está disponível no site de Demografia UFRN (s. d.). O algoritmo desenvolvido facilita a aplicação do método sempre que houver uma atualização ou mudanças nos dados de população para as áreas de interesse.

Dados e métodos

Os dados básicos para qualquer aplicação do método proposto neste artigo são: *a)* população por sexo e grupos quinquenais de idade (0 a 4, 5 a 9,..., 90+) para todas as áreas menores de interesse e para a área maior que compreende geograficamente as áreas menores nos dois últimos censos demográficos; *b)* população projetada para a área maior, por sexo e grupos quinquenais de idade, em todos os anos calendários quinquenais, ao longo de todo o período de projeção; *c)* taxa total e taxas específicas de fecundidade da área maior, projetadas por grupos quinquenais de idade das mulheres ao longo do período reprodutivo para todos os anos calendários quinquenais ao longo de todo o período de projeção, e *d)* tábuas de mortalidade projetadas para as áreas maiores, por grupos quinquenais de idade e sexo, para todos os anos calendários quinquenais, ao longo de todo o período de projeção.

Os dados censitários e as estimativas e projeções de taxas demográficas para as áreas maiores, como estados, departamentos, etc., são comumente disponibilizados por institutos ou agências de estatísticas oficiais. Requer-se, no caso do método aqui proposto, que o intervalo de tempo entre os dois últimos censos seja de 10 anos. Entretanto, com poucas alterações no algoritmo, o método pode ser aplicado quando o intervalo de tempo entre os dois últimos censos for de 5 anos. Para outros intervalos intercensitários, sugere-se a interpolação das populações por sexo e idade no tempo via modelos geométricos ou exponenciais de crescimento populacional.

Para os resultados aqui apresentados foram utilizados os dados de população por sexo e grupos quinquenais de idade dos 5.570 municípios e das 26 unidades federativas² do Brasil, nos censos demográficos de 1991, 2000 e 2010. As estimativas e projeções de taxas demográficas e população, para todas as unidades federativas do Brasil são disponibilizadas pelo Instituto Brasileiro de Geografia e Estatística, cuja última revisão foi realizada em 2018 (IBGE, 2018a).

O método de relação das coortes

O *método de relação das coortes*, proposto por Duchesne (1989), poderia ser considerado com uma variante do método de coortes componentes (González e Torres, 2012) que requer, porém, dados mais fáceis de serem obtidos para áreas menores em muitos países ou regiões. A lógica demográfica do método está fundamentada em uma medida comumente denominada *razão de mudança de coortes* ou *razão de sobrevivência intercensitária* (Hamilton e Perry, 1962; Swanson e Tedrow, 2013), que representa a mudança da população na coorte etária entre os dois censos, comumente decenais, e pode ser específica para cada sexo.

O método pressupõe a disponibilidade de projeções por grupos etários e períodos de tempo quinquenais para as áreas maiores, preferencialmente realizadas pelo método demográfico das coortes componentes (Granados, 1989). Ademais, o método baseia-se em dois parâmetros definidos mediante razões de sobrevivência de coortes quinquenais: *a)* coeficiente de crescimento das coortes etárias da área maior, e *b)* fator diferencial de crescimento das coortes etárias das áreas menores em relação às respectivas coortes etárias da área maior. De um modo geral, com algumas

² O Distrito Federal é uma unidade federativa que compõe uma única área sem subáreas municipais. Portanto, para este caso as projeções são exatamente aquelas disponibilizadas pelo IBGE, na revisão de 2018.

adaptações para o primeiro e último grupos etários, o algoritmo de projeção baseia-se na seguinte equação:

$${}_5P_{x+5,i}^{t+5} = {}_5P_{x,i}^t \times {}_5RS_x^{t,t+5} \times {}_5K_{x,i} \quad (1)$$

onde cada parâmetro pode ser interpretado como segue:

${}_5P_{x+5,i}^{t+5}$: população da área menor i no grupo etário $x+5, x+10$ projetada para o ano $t+5$;

${}_5P_{x,i}^t$: população da área menor i no grupo etário $x, x+5$ recenseada no ano t ;

${}_5RS_x^{t,t+5} = \frac{{}_5P_{x+5,i}^{t+5}}{{}_5P_x^t}$: coeficiente de crescimento ou razão de sobrevivência correspondente à população da área maior no grupo etário $x, x+5$ que alcança o grupo etário $x+5, x+10$ entre os anos t e $t+5$. Esse coeficiente pressupõe projeções já realizadas para a população da área maior por grupos etários quinquenais entre t e $t+5$.

O último parâmetro do lado direito da equação 1 trata-se do fator diferencial de crescimento das coortes etárias nas áreas menores em relação às respectivas coortes etárias na área maior. Este fator deve ser calculado com base nos dois últimos censos disponíveis e deve ser mantido constante em todos os períodos quinquenais de projeção, motivo pelo qual este fator não é indexado pelo período de tempo na equação 1. Sendo considerado como um dos principais parâmetros da equação 1, este fator K carece de maior detalhamento. Segundo Duchesne (1989), o cálculo do fator K para cada área menor i é realizado como segue:

$${}_5K_{x,i} = \frac{\frac{{}_5P_{x+5,i}^{t+5}}{{}_5P_{x,i}^t}}{{}_5P_{x+5,i}^{t+5} / \frac{{}_5P_x^t}}{{}_5P_{x+5,i}^{t+5}}} \quad (2)$$

onde:

$\frac{{}_5P_{x+5,i}^{t+5}}{{}_5P_{x,i}^t}$: é a razão de sobrevivência intercensitária da população da área menor i no grupo etário $x, x+5$ no primeiro censo (ano t) que alcança o grupo etário $x+5, x+10$ no segundo censo (ano $t+5$). A razão no denominador é a medida equivalente para a área maior.

A definição de K tal como expressa na equação 2 pressupõe que o intervalo intercensitário entre os dois últimos censos disponíveis seja de 5 anos, o que dificilmente ocorre em muitos países, como no caso do Brasil. Duchesne (1989), sugeriu um ajuste no fator K com o objetivo de incorporar o efeito do diferencial de crescimento para um intervalo intercensitário de 5 anos, ainda que os dois últimos censos demográficos sejam decenais. Assim, com base nas populações observadas em censos decenais, um fator K para uma área menor i , para grupos etários e períodos de tempo quinquenais, pode ser estimado por:

$${}_5K_{x,i} = 0,5 \left[\left(\frac{\frac{{}_5P_{x+5,i}^{t+10}}{{}_5P_{x-5,i}^t}}{{}_5P_{x+5,i}^{t+10} / \frac{{}_5P_x^t}}{{}_5P_{x+5,i}^{t+10}}} \right)^{0,5} + \left(\frac{\frac{{}_5P_{x+10,i}^{t+10}}{{}_5P_{x,i}^t}}{{}_5P_{x+10,i}^{t+10} / \frac{{}_5P_x^t}}{{}_5P_{x+10,i}^{t+10}}} \right)^{0,5} \right] \quad (3)$$

Como exposto pelas equações 2 e 3, uma vez que o crescimento de uma coorte específica numa área menor qualquer pode ser maior ou menor que o crescimento da respectiva coorte na área maior, os valores de K por grupos de idade devem variar ao redor de uma unidade. Um fator K menor que 1 significa que o crescimento da coorte na área menor é menor que o crescimento da respectiva coorte na área maior. Por outro lado, um fator K maior que 1 significa que o crescimento da coorte na área menor é maior que o crescimento da respectiva coorte na área maior. Assim, pode-se dizer que a RS da área maior na equação 1 é ajustada pelo fator K da área menor i .

Uma das vantagens do método de relação das coortes é a pouca exigência de dados para sua aplicação. Como observado nas equações anteriores, tanto para o cálculo do fator K quanto para implementação das projeções em áreas menores pela equação 1, os dados necessários são: população recenseada por grupos quinquenais de idade e sexo das áreas menores e na área maior nos dois últimos censos; populações projetadas por sexo, intervalos etários e período de tempo quinquenais para as áreas maiores. Adicionalmente, para determinação dos nascimentos nas áreas menores em cada quinquênio de projeção, são necessárias as taxas de fecundidade e tábuas de mortalidade projetadas para as áreas maiores. Estas últimas informações são insumos para aplicação do método das coortes componentes para projeção das áreas maiores e são, portanto, comumente disponibilizadas pelas agências oficiais de estatísticas nacionais.

Talvez a principal desvantagem do método seja a constância do fator nas projeções populacionais das áreas menores pela equação 1. Qualquer mudança temporária e considerável em qualquer das componentes demográficas de uma área menor, entre os dois últimos censos, implicaria em grandes mudanças na estrutura etária da população projetada para aquela área menor. Neste artigo, lidamos com esta limitação ajustando o cálculo de K por um estimador bayesiano empírico que, além de minimizar a variabilidade nos valores de K produz uma suavização espacial estabilizando estes valores em áreas com baixo contingente populacional.

Implementação do método de relação de coortes

A implementação da projeção populacional por sexo grupos quinquenais de idade para áreas menores, mediante aplicação da equação 1, requer alguns ajustes para o último grupo etário (de intervalo aberto à direita), e também para o primeiro grupo etário (0 a 4 anos). Para o último intervalo etário a equação de projeção é como segue:

$${}_wP_{x+5,i}^{t+5} = {}_wP_{x,i}^t \times {}_wRS_x^{t,t+5} \times {}_wK_{x,i} \quad (4)$$

onde cada parâmetro pode ser interpretado como segue:

${}_wP_{x+5,i}^{t+5}$: população da área menor i no grupo etário $x+5$, $x+5+w$ projetada para o ano $t+5$;

${}_wP_{x,i}^t$: população da área menor i no grupo etário x , $x+w$ no ano t ;

${}_wRS_x^{t,t+5} = \frac{{}_wP_{x+5}^{t+5}}{{}_wP_x^t}$: coeficiente de crescimento ou razão de sobrevivência correspondente à população da área maior no grupo etário x , $x+w$ que alcança o grupo etário $x+5$, $x+5+w$ entre os anos t e $t+5$.

Numa perspectiva de projeção populacional por intervalo de tempo quinquenal, a população no primeiro grupo quinquenal de idade se refere aos nascimentos

sobreviventes ao longo do quinquênio e, portanto, não pode ser obtida pelo algoritmo da equação 1. Ademais, a baixa qualidade dos dados de população e registros vitais (nascimentos e óbitos) em áreas menores, em particular no Brasil, impossibilitam o cálculo das taxas específicas de nascimentos e óbitos, de modo que tais taxas possam ser utilizadas para estimar os nascimentos sobreviventes no quinquênio. A solução apontada por Duchesne (1989) é a utilização de um índice diferencial de fecundidade, entre cada uma das áreas menores e a área maior, calculado com base nos dados censitários. O cálculo deste índice diferencial de fecundidade requer, inicialmente, estimativas da Razão Criança-Mulher (Lee, 1957), tanto para as áreas menores quanto para a área maior. Assim, de acordo com Duchesne (1989), o algoritmo detalhado a seguir busca suprir tais limitações nos dados e possibilita estimar indiretamente as quantidades de nascimentos sobreviventes nas áreas menores durante um intervalo de tempo quinquenal.

Primeiramente, para cada área menor i no último censo, calcula-se a Razão Criança-Mulher:

$$RCM_i = \frac{{}_5P_{0,i}}{30P_{15,i}^f} \quad (5)$$

onde o sobrescrito “ f ” no denominador faz referência à população de mulheres no intervalo etário reprodutivo, assumindo que este intervalo se inicia no grupo etário 15 a 19 e termina aos 45 a 49 anos, ou seja, as mulheres expostas ao risco de gerar um nascimento. ${}_5P_{0,i}$ é o número de crianças de 0 a 4 anos completos na pequena área i no último censo.

Em seguida, para a área maior no último censo, também se calcula a Razão Criança-Mulher:

$$RCM = \frac{{}_5P_0}{30P_{15}^f} \quad (6)$$

onde, similar à equação 5, o sobrescrito “ f ” no denominador também faz referência à população de mulheres no intervalo etário reprodutivo de 15 a 45 anos (desta vez, na área maior). Da mesma forma, o ${}_5P_0$ é o número de crianças de 0 a 4 anos completos, agora da área maior.

Uma vez determinadas as razões criança-mulher nas áreas menores e na área maior, calcula-se o seguinte índice diferencial de fecundidade para cada área menor i :

$$IDF_i = \frac{RCM_i}{RCM} \quad (7)$$

Posteriormente, calcula-se a *taxa de fecundidade total* para cada área menor i equivalente ao quinquênio de projeção $t, t+5$:

$$TFT_i^{t,t+5} = IDF_i(TFT^{t,t+5}) \quad (8)$$

Observa-se que a taxa de fecundidade total na área menor i depende do índice diferencial de fecundidade de cada área menor e da taxa de fecundidade total projetada na área maior. De posse das taxas de fecundidade totais, calculam-se as taxas específicas de fecundidade (TEF) para cada área menor i :

$${}_5TEF_{x,i}^{t,t+5} = \left(\frac{{}_5TEF_x^{t,t+5}}{TFT^{t,t+5}} \right) TFT_i^{t,t+5} \quad (9)$$

A razão entre a TEF e a TFT no período de projeção $t, t+5$, na área maior, representa o percentual da TEF das mulheres no grupo etário $x, x+5$, em relação à TFT no mesmo grupo etário. As TEF para a área menor i determinarão o volume de nascimentos desta localidade no período de projeção $t, t+5$.

Antes de proceder ao cálculo dos nascimentos na área menor i , determina-se o volume de mulheres expostas ao risco de gerar um nascimento, que pode ser aproximado pelo total de mulheres no meio do intervalo de tempo de 5 anos:

$${}_5\bar{P}_{x,i}^{f;t,t+5} = \frac{1}{2} ({}_5P_{x,i}^{f,t} + {}_5P_{x,i}^{f,t+5}) \quad (10)$$

Lembrando que as populações ${}_5P_{x,i}^{f,t}$ e ${}_5P_{x,i}^{f,t+5}$ foram obtidas pela aplicação da equação 1 para o sexo feminino. De posse do número de mulheres no meio do período quinquenal de projeção, e também das TEF aplicáveis a todo quinquênio para cada área menor i , determina-se o número de nascimentos (B) em cada localidade i entre t e $t+5$:

$$B_i^{t,t+5} = \sum_{x=15}^{45} [{}_5\bar{P}_{x,i}^{f;t,t+5} ({}_5TEF_{x,i}^{t,t+5})] 5 \quad (11)$$

onde a multiplicação por 5, em cada termo do somatório, deve-se ao fato de que as TEF representam o risco médio anual de nascimentos dentro do grupo etário $x, x+5$. Para estimar os nascimentos em todo o intervalo etário $x, x+5$ é necessário multiplicar as TEF por 5.

Os nascimentos determinados pela equação 11 estão expostos a um risco de morte desde o nascimento até antes de completarem o 5.º aniversário. Ademais, a equação 11 não distingue entre nascimentos femininos e masculinos, informação importante para determinar o total de mulheres que, no futuro estarão expostas ao risco de gerar um nascimento na área menor i . Assim, o cálculo dos nascimentos sobreviventes masculinos e femininos, respectivamente, na área menor i dar-se-á por:

$${}_5P_{0,i}^{m,t+5} = B_i^{t,t+5} \left(\frac{RSN}{1+RSN} \right) \left(\frac{{}_5L_0^m}{{}_5I_0^m} \right)^{t,t+5} K_{b,i}^m \quad (12)$$

e

$${}_5P_{0,i}^{f,t+5} = B_i^{t,t+5} \left(\frac{1}{1+RSN} \right) \left(\frac{{}_5L_0^f}{{}_5I_0^f} \right)^{t,t+5} K_{b,i}^f \quad (13)$$

onde os fatores $RSN/(1+RSN)$ e $1/(1+RSN)$ determinam as parcelas de nascimentos masculinos e femininos, respectivamente. Já a razão ${}_5L_0/{}_5I_0$ para cada sexo, determina o risco de sobrevivência ao nascimento na área maior e deve ser obtido com base nas tábuas de vida projetadas para a área maior. Os termos ${}_5P_{0,i}^{m;t,t+5}$ e ${}_5P_{0,i}^{f;t,t+5}$ são as populações de 0 a 4 anos completos da área menor i projetadas para o período entre t e $t+5$.

As quantidades $K_{b,i}^m$ e $K_{b,i}^f$ em ambas as equações 12 e 13 representam os fatores diferenciais de crescimento ao nascer, para nascimentos masculinos e femininos respectivamente, para cada área menor i e podem ser obtidas por:

$$K_{b,i}^m = ({}_5K_{0,i}^m)^{0,5} \quad (14)$$

$$K_{b,i}^f = ({}_5K_{0,i}^f)^{0,5} \quad (15)$$

onde ${}_5K_0$, seja masculino ou feminino, é equivalente ao período de um quinquênio e, por isso, precisa ser ajustado para representar um K equivalente a um ano, já que o índice diferencial de crescimento ao nascimento refere-se somente ao primeiro ano.

Duchesne (1989) destaca algumas observações importantes sobre o cálculo do índice diferencial de crescimento ao nascimento da área menor em relação à área maior entre os dois últimos censos populacionais (tal como definido nas equações 14 e 15). Primeiro, se o intervalo intercensitário é de 10 anos, não temos informações necessárias para estimar o fator ${}_5K_0$, tal como expresso pela equação 3. Neste caso, recomenda-se a utilização do fator ${}_5K_5$, este sim podendo ser obtido pela equação 3. Segundo, nos casos em que a migração nas primeiras idades é intensa (quando as mães migram com seus filhos ou geram nascimentos no local de destino logo após a migração), sugere-se utilizar, no cálculo do fator K_b , os fatores ${}_5K_{20}$ ou ${}_5K_{25}$ ao invés de ${}_5K_5$. Nesse artigo, usamos o ${}_5K_5$.

Um estimador bayesiano empírico para o ajuste do fator K

O fator diferencial de crescimento, o fator K , pode ter alta variabilidade, principalmente nas áreas menores e nos grupos etários de menor população. Neste sentido, sugerimos uma alternativa metodológica para minimizar as variâncias dos fatores K_i a partir de um estimador contração, da classe bayesiano empírico. O fator K pode ser determinado como segue:

$${}_5K_{x,i} = 0,5 \left[\left(\frac{{}_5P_{x+5,i}^{t+10}}{ESP_{i,1}} \right)^{0,5} + \left(\frac{{}_5P_{x+10,i}^{t+10}}{ESP_{i,2}} \right)^{0,5} \right] \quad (16)$$

Podemos ainda simplificar e escrever como na fórmula 17 a seguir. Os estimadores bayesianos empíricos propostos são para estimar $K_{i,1}$ e $K_{i,2}$ da fórmula 17.

$${}_5K_{x,i} = 0,5 \left[(K_{i,1})^{0,5} + (K_{i,2})^{0,5} \right] \quad (17)$$

onde,

$$ESP_{i,1} = {}_5P_{x-5,i}^t \frac{{}_5P_{x+5,i}^{t+10}}{{}_5P_{x-5,i}^t}; ESP_{i,2} = {}_5P_{x,i}^t \frac{{}_5P_{x+10,i}^{t+10}}{{}_5P_{x,i}^t}; K_{i,1} = \frac{{}_5P_{x+5,i}^{t+10}}{ESP_{i,1}} \text{ e } K_{i,2} = \frac{{}_5P_{x+10,i}^{t+10}}{ESP_{i,2}} \quad (18)$$

$ESP_{i,1}$ é interpretado como a população esperada na área menor i no grupo etário $x+5$, $x+10$ anos, no ano $t+10$, sob a hipótese de que a razão de sobrevivência da coorte na área menor i seja a mesma da área maior no mesmo período. A interpretação para $ESP_{i,2}$ é

análoga, $ESP_{i,2}$ é a população esperada na área menor i no grupo etário $x+10$; $x+15$ anos, no ano $t+10$, sob a hipótese descrita acima. Na equação 16, tanto ${}_5P_{x+5,i}^{t+10}$ quanto ${}_5P_{x+10,i}^{t+10}$ são contagens e podemos tratá-los, estocasticamente, como uma variável aleatória com distribuição de Poisson. A descrição a seguir se refere ao primeiro termo (${}_5P_{x+5,i}^{t+10}$) sendo análogo ao segundo termo:

$${}_5P_{x+5,i}^{t+10} \sim \text{poisson}(ESP_{i,1} \cdot \theta_i) \quad (19)$$

onde θ_i é o risco da população da coorte ser maior ou menor que o esperado sob a hipótese de que a área menor i possua a mesma razão de sobrevivência que a área maior. Com isso, tem-se que a média condicional de ${}_5P_{x+5,i}^{t+10}$ é $E({}_5P_{x+5,i}^{t+10}/\theta_i) = ESP_{i,1}\theta_i$. O estimador de máxima-verossimilhança para θ_i é o fator $K_{i,1}$, que pode ser escrito pela equação 18.

Seguindo o arcabouço teórico de Marshall (1991), θ_i tem distribuição *a priori* não especificada, com momentos $E(\theta_i)=m$ e $v(\theta_i)=A$. O estimador bayesiano empírico para $K_{i,1}$ é dado por:

$$\hat{\theta}_i = \hat{m} + \hat{c}_i(K_{i,1} - \hat{m}) \quad (20)$$

onde,

$$\hat{m} = \frac{\sum_{i=1}^n {}_5P_{x+5,i}^{t+10}}{\sum_{i=1}^n ESP_{i,1}} \quad (21)$$

Na equação 21, n indica o número de pequenas áreas e \hat{m} deve ser interpretado como sendo o fator K médio para todo o conjunto de áreas, ou seja, a área maior. Portanto, \hat{m} é 1. Ainda segundo Marshall (1991), o fator \hat{c}_i na equação 20, representa o fator de contração do estimador bayesiano empírico, que é dado por:

$$\hat{c}_i = \frac{v(\theta_i)}{v(K_{i,1})} = \frac{\hat{A}}{\hat{A} + \frac{\hat{m}}{ESP_{i,1}}} \quad (22)$$

onde,

$\hat{A} = s^2 - \hat{m} \left[\sum_{i=1}^n \frac{{}_5P_{x+5,i}^{t+10}/T}{ESP_{i,1}} \right]$ é o estimador para a variância do parâmetro θ_i . O denominador da expressão 22 é o estimador para a variância de $K_{i,1}$, já $s^2 = \frac{\sum_{i=1}^n {}_5P_{x+5,i}^{t+10}(K_{i,1} - \hat{m})^2}{T}$, é a variância de $K_{i,1}$, ponderada pelo tamanho do grupo etário na pequena área i e $T = (\sum_{i=1}^n {}_5P_{x+5,i}^{t+10})$. É importante lembrar que os valores populacionais da fórmula 21 e da legenda da fórmula 22, expressos aqui neste parágrafo, se referem à população do primeiro dos dois censos utilizados como base de entrada para o cálculo dos fatores K . No artigo original de Marshall (1991) o estimador bayesiano empírico é uma taxa de incidência, diferente do que se propõe neste trabalho. Ao longo do Apêndice deste artigo encontram-se as demonstrações detalhadas para $E(K_{i,1})$, $v(K_{i,1})$ e \hat{A} , respectivamente.

Em suma, \hat{c}_i é o quociente entre a variância de θ_i e a variância de seu estimador, dado por $K_{i,1}$ e varia entre 0 e 1. A lógica é que quando a variância de $K_{i,1}$ é muito alta, portando um estimador pouco confiável, \hat{c}_i tenderá à 0. Nesses casos, o valor de θ_i (que é o estimador bayesiano empírico para $K_{i,1}$) na equação 20 tenderá para a média (\hat{m}) dos seus vizinhos ou, como utilizado neste trabalho, a média da Unidade da Federação. Por isso que chamamos de estimador contração, pois a suavização operada por este estimador da equação 20 reduz ou contrai o valor inicial da estimativa de $K_{i,1}$ para o

valor médio. Nesse caso, se por hipótese o valor de \hat{c}_i de uma determinada pequena área fosse 1, seu fator $K_{i,1}$, originalmente calculado, seria descartado pelo fator $K_{i,1}$ médio da Unidade da Federação. Estaríamos, portanto, assumindo que aquela pequena área i teria exatamente a mesma relação de crescimento da área maior.

Por outro lado, se a variância de $K_{i,1}$ é muito pequena e tender a 0, o valor de \hat{c}_i tenderá para 1. Pela equação 20 percebe-se que, quanto mais próximo \hat{c}_i for de 1 menor será a contração, redução ou suavização operada no $K_{i,1}$ originalmente calculado para a área i , o que faz sentido na medida em que se a variância de $K_{i,1}$ é baixa este já é um estimador confiável, não sendo necessário grande ajuste na estimativa inicial.

Também podemos entender θ_i como um estimador de caráter espacial, uma vez que contrai a estimativa inicial de $K_{i,1}$ da área menor para o valor médio, e essa média pode ser dos vizinhos, dos similares segundo algum critério ou de uma área maior que contém a pequena área i , como utilizado neste trabalho.

Por fim, essa abordagem é da classe bayesiana uma vez que adotamos o arcabouço teórico bayesiano onde os parâmetros não são fixos, tendo distribuição de probabilidade associada, ainda que nem sempre especificadas como aqui. Desta forma, nesta abordagem assumimos uma abordagem estocástica para θ_i na medida em que assumimos variabilidade e a estimamos. Mas não estamos gerando distribuição à posteriori, uma vez que a priori para θ_i não foi especificada. Como dito anteriormente, todo esse desenvolvimento teórico é análogo para o $K_{i,2}$ da fórmula 17. No anexo 4 estão todas as fórmulas para o estimador bayesiano empírico para o $K_{i,2}$.

Resultados

O método de relação das coortes, com ajuste pelo estimador bayesiano empírico no cálculo do fator diferencial de crescimento, foi aplicado para projeções populacionais dos 5.570 municípios do Brasil entre 2010 e 2030, por sexo e grupos quinquenais de idade (0-4, 5-9, 10-14,..., 90+). As áreas maiores foram as 26 unidades federativas. O Distrito Federal não contém delimitações municipais e, por isso, não foi incluído nas projeções de áreas menores.

Inicialmente, para avaliação dos resultados com o ajuste pelo estimador bayesiano empírico proposto neste artigo, aplicou-se o método para projeções populacionais em todos os municípios do Brasil entre 2000 e 2010 não envolvidos em processo de emancipação.³ Uma avaliação dos resultados de uma projeção populacional para um ano específico só é possível quando se conhece a população observada no mesmo ano. Dessa forma, utilizou-se os dados de população dos municípios nos censos de 1991 e 2000 para projetar as populações municipais para o ano de 2010 e, então, comparar as estimativas com a população observada no censo de 2010.

Primeiramente, os fatores diferenciais de crescimento (K) foram determinados sem qualquer tipo de ajuste ou suavização. Em seguida, utilizou-se o estimador bayesiano empírico para incorporar uma suavização espacial no cálculo dos fatores K . Então, projeções populacionais pelo método de relação de coortes foram realizadas considerando-se essas duas formas distintas de obtenção dos fatores diferenciais de crescimento.

3 Emancipação municipal pode ocorrer no Brasil de diversas formas. Um município pode ser desmembrando em dois ou três ou determinado município pode ser criado com base em populações cedidas por dois ou mais municípios distintos.

Avaliação do erro de projeção com e sem o ajuste no fator K

A precisão da estimativa populacional para um intervalo de 10 anos foi avaliada considerando-se o erro percentual (EP) que é o quociente da diferença entre as populações observada e estimada para 2010:

$$EP = \frac{p_{2010}^{censo} - p_{2010}^{projetada}}{p_{2010}^{censo}} \quad (23)$$

Com base no EP calculou-se o Erro Percentual Absoluto (EPA), que fornece a magnitude do EP . Por fim, estimou-se o Erro Percentual Absoluto Mediano ($EPAM$), que é simplesmente a mediana do EPA para o total ou um conjunto de áreas de interesse. Na Tabela 1 são apresentados os Erros Percentuais Absolutos Medianos nas projeções populacionais para o ano de 2010, por sexo e grupos quinquenais de idade, considerando as duas formas distintas de cálculo dos fatores diferenciais de crescimento (K). Observa-se maior redução no $EPAM$ para ambos os sexos quando o fator K suavizado pelo estimador bayesiano empírico (BE) é utilizado. Embora seja uma redução modesta no $EPAM$ pode-se concluir que a metodologia proposta melhora a eficácia do método proposto por Duchesne (1989) em termos da projeção para a população total das localidades menores.

Tabela 1
Erro percentual absoluto mediano para as projeções populacionais municipais por sexo. Brasil, 2000-2010 (em porcentagens)

Cálculos alternativos do fator de crescimento	Homens	Mulheres
K original	11,40	9,49
K suavizado por BE	10,20	8,52

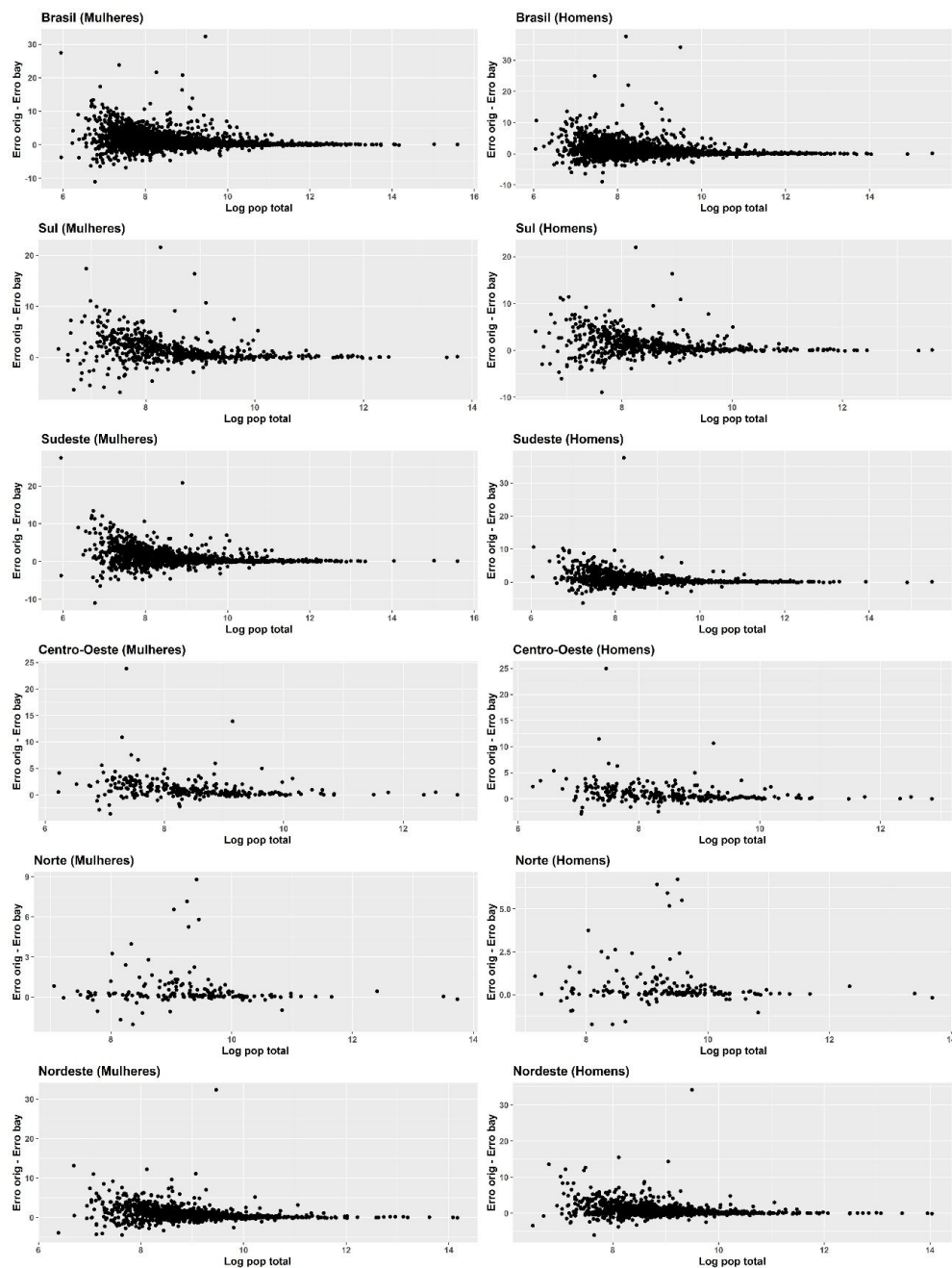
Nota: BE (bayesiano empírico).

Fonte: IBGE (1991, 2000, 2010).

Uma análise do diferencial dos erros de projeção (com K original e com K ajustado pelo bayesiano empírico) para todos os municípios do Brasil não envolvidos em emancipação, por grande região e por tamanho populacional é apresentada na Figura 1.

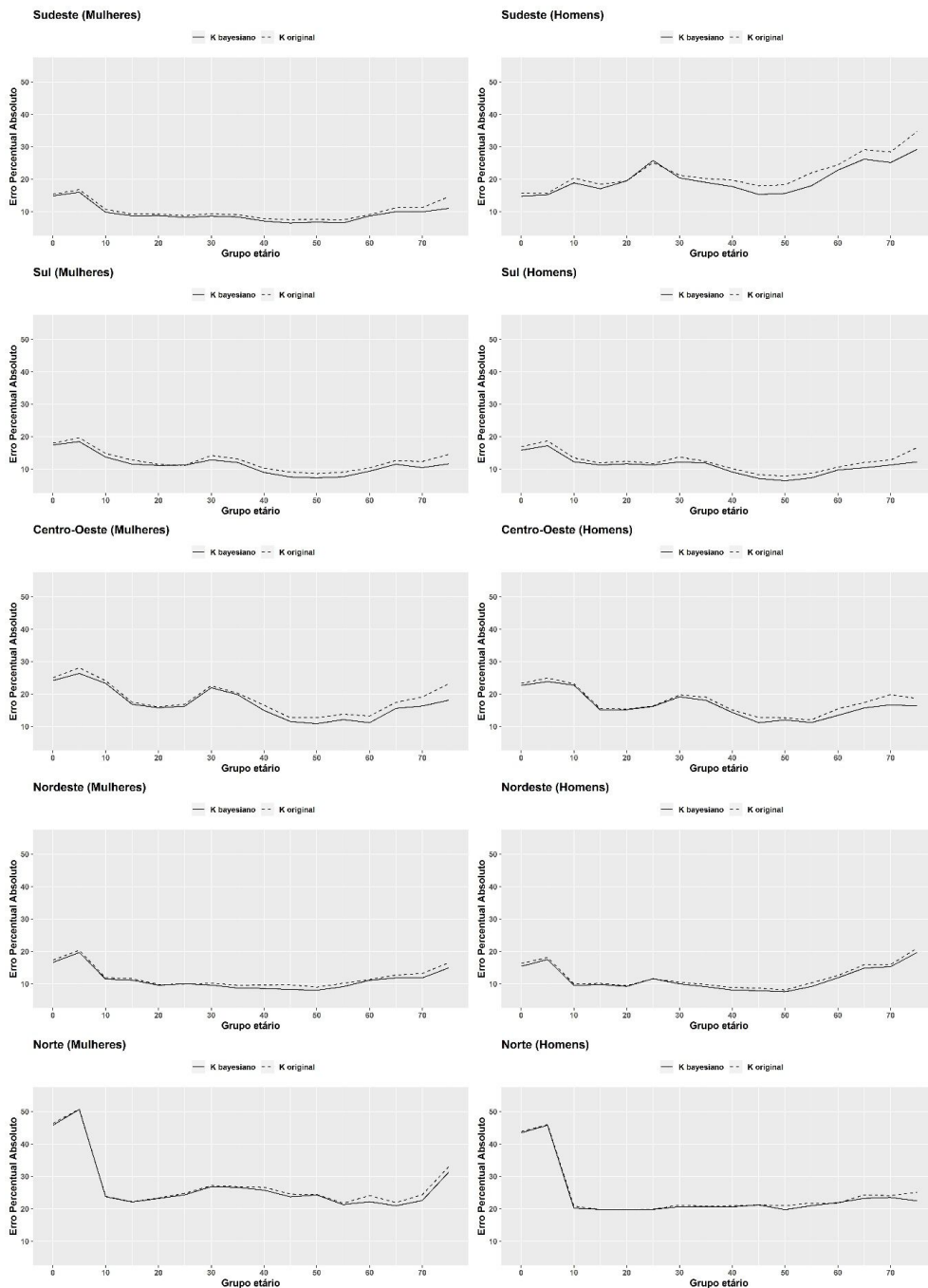
Seja para o Brasil como um todo ou para grandes regiões, os diagramas de dispersão na Figura 1 mostram a relação entre as diferenças dos erros percentuais absolutos de projeção e os tamanhos populacionais dos municípios. Observa-se, tanto para o Brasil quanto para suas 5 grandes regiões, que as diferenças entre os erros (erro com o K original menos erro com K pelo estimador bayesiano empírico) são predominantemente maiores que zero, evidenciando que a redução dos erros com ajuste do fator K pelo estimador bayesiano empírico é generalizada em todas as regiões do país. Mais que isso, as relações apresentadas na Figura 1 mostram que quanto menor o tamanho da população, maiores são as diferenças entre os erros de projeção. Este resultado ressalta uma contribuição importante do ajuste pelo estimador bayesiano nas estimativas do fator K para projeções populacionais pelo método de relação das coortes, que é redução do erro de projeção em áreas com menores tamanhos populacionais.

Figura 1
Diferença entre os erros percentuais absolutos em relação ao tamanho da população municipal por sexo. Brasil e Grandes Regiões, 2000-2010



Fonte: IBGE (1991, 2000, 2010).

Figura 2
Erro percentual absoluto mediano, por grupos de idade, sexo e grandes regiões. Brasil, 2000-2010



Fonte: IBGE (1991, 2000, 2010).

Essa melhor precisão nas projeções populacionais de áreas menores pelo método de Relação de Coortes, quando ajustado pelo estimador BE, também é evidente quando comparamos os erros de projeção por sexo e grupos de idade. A Figura 2 mostra os erros percentuais absolutos medianos das projeções populacionais municipais por grandes regiões, sexo e grupos de idade (0-4, 5-9,...,75+) mediante o uso do fator K original ou ajustado pelo estimador bayesiano empírico (BE). O padrão etário dos erros é muito parecido entre as duas formas distintas de cálculo do fator K . A suavização do fator K pelo estimador BE reduziu o erro de projeção em quase todos os grupos etários para ambos os sexos, especialmente acima dos 30 anos. Embora os ganhos sejam tímidos, as precisões melhoraram em todas as regiões do Brasil.

Projeções populacionais para os 5.570 municípios do Brasil entre 2010 e 2030

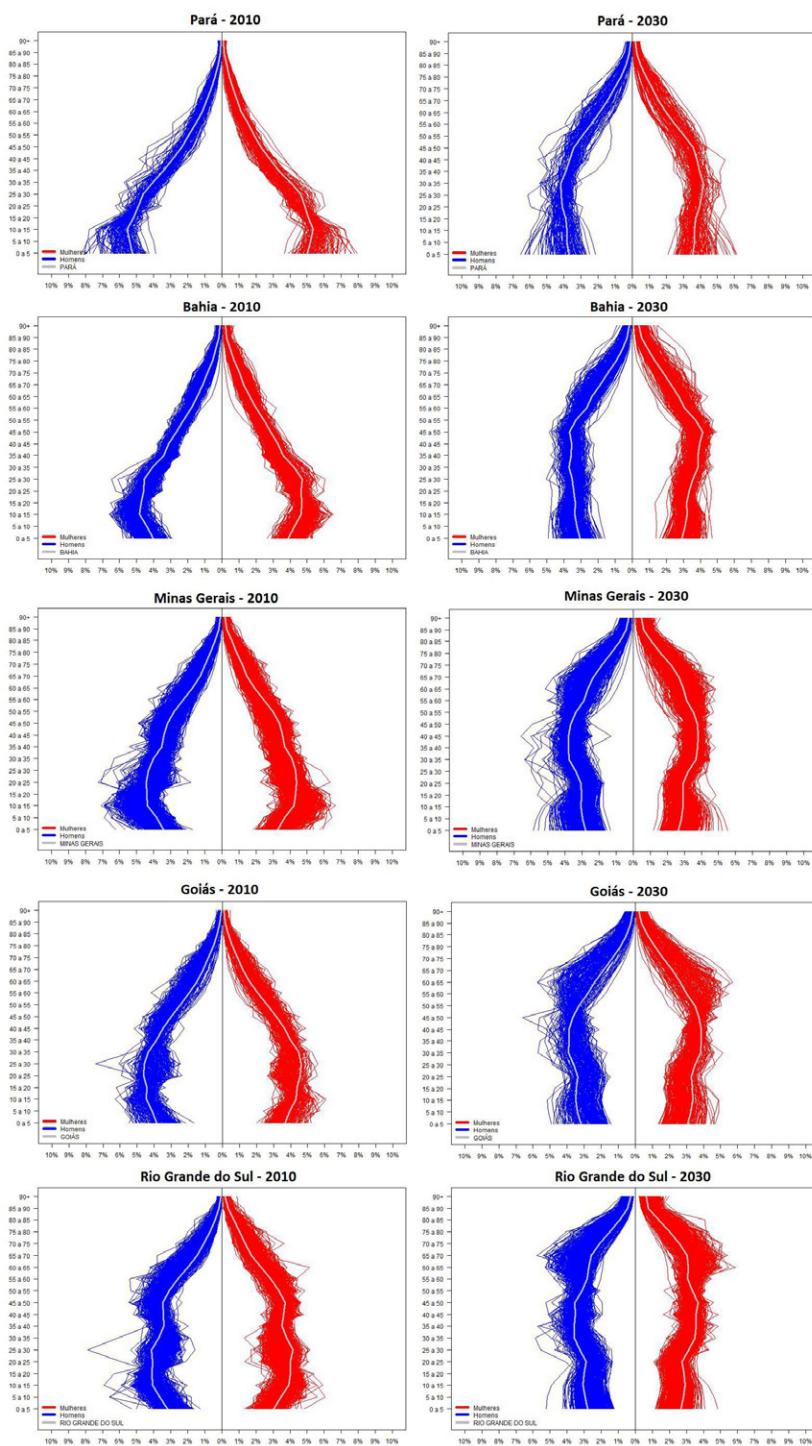
Utilizando o método de relação de coortes (Duchesne, 1989) com suavização dos fatores diferenciais de crescimento pelo estimador bayesiano empírico, foram realizadas projeções populacionais por sexo, grupos de idade e períodos tempo quinquenais para os 5.570 municípios brasileiros no período de 2010 a 2030. A população Brasileira tem experimentado profundas transformações em sua estrutura etária desde metade do século passado. A queda vertiginosa nas taxas de fecundidade em todas as regiões sem perspectivas de reversão do declínio dessas taxas e as tendências de queda continuada nas taxas de mortalidade infantil apontam para um cenário onde a razão de dependência continuará aumentando em todas as regiões do país com importantes impactos nas áreas de saúde, previdência e sociais.

Os resultados das projeções municipais neste estudo não contradizem as tendências nas estruturas dos estados ou grandes regiões. Como exemplo, na Figura 3 mostramos a evolução nas estruturas etárias municipais para 5 estados brasileiros selecionados, um de cada região, entre 2010 e 2030. Nesta figura as linhas azuis e vermelhas representam as estruturas etárias masculinas e femininas, respectivamente. Já as linhas cinzas centralizadas, representam as estruturas etárias para o respectivo estado. Nota-se que a tendência na estrutura etária de cada estado representa uma média das tendências municipais. Percebe-se também que, as áreas menores nas regiões sul/sudeste apresentam, já em 2010, em média, estruturas etárias mais envelhecidas que as áreas menores das regiões norte/nordeste. As pirâmides demonstram o rápido aumento das proporções de população acima dos 65 anos e diminuição das populações mais jovens, com estreitamento da base das pirâmides, fenômeno observado em todas as unidades federativas brasileiras.⁴

As Figuras 4 e 5 mostram a evolução de alguns indicadores em consequência das mudanças na estrutura etária: aumento do percentual da população acima dos 65 anos; evolução da razão de dependência total, com ponto de inflexão em 2020, tanto para homens quanto para mulheres; aumento da razão de dependência idosa (população com 65 anos e mais dividido pela população entre 15 e 64 anos) e; diminuição da razão de dependência jovem (população até 14 anos dividido pela população entre 15 e 64 anos). Cabe ressaltar que os resultados demonstrados nas Figuras 3 e 4 são a média municipal dos 5.570 municípios brasileiros. Neste sentido, quando se fala que em 2020 a razão de dependência total deixará de declinar para começar a aumentar, estamos dizendo que em média, as razões de dependência municipal no Brasil passarão a aumentar a partir de 2020. Essa explicação é análoga para os outros 3 indicadores.

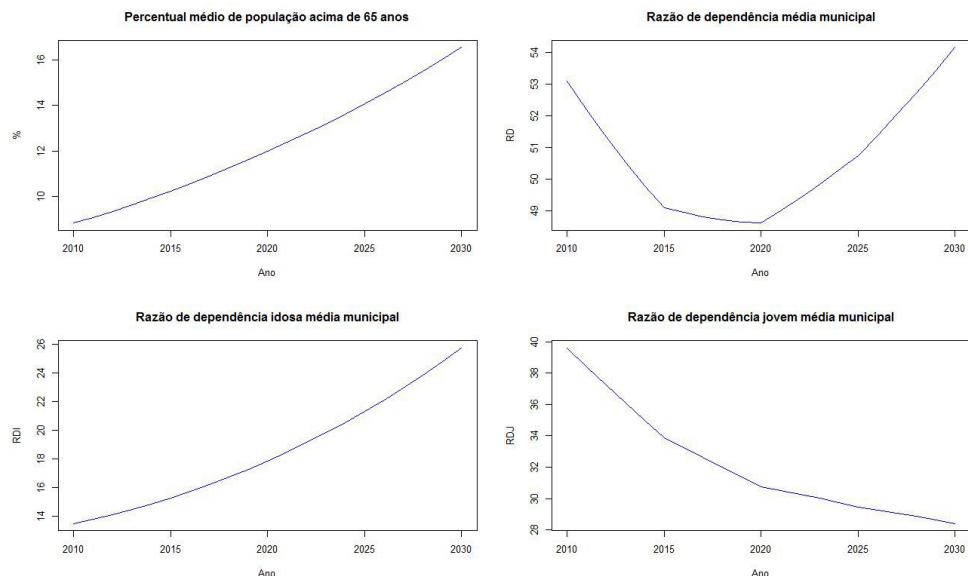
4 As demais pirâmides etárias e rotinas computacionais utilizadas neste trabalho estarão disponíveis no site de Demografia UFRN (s. d.).

Figura 3
 Pirâmides etárias municipais para 5 estados do Brasil, 2010 e 2030



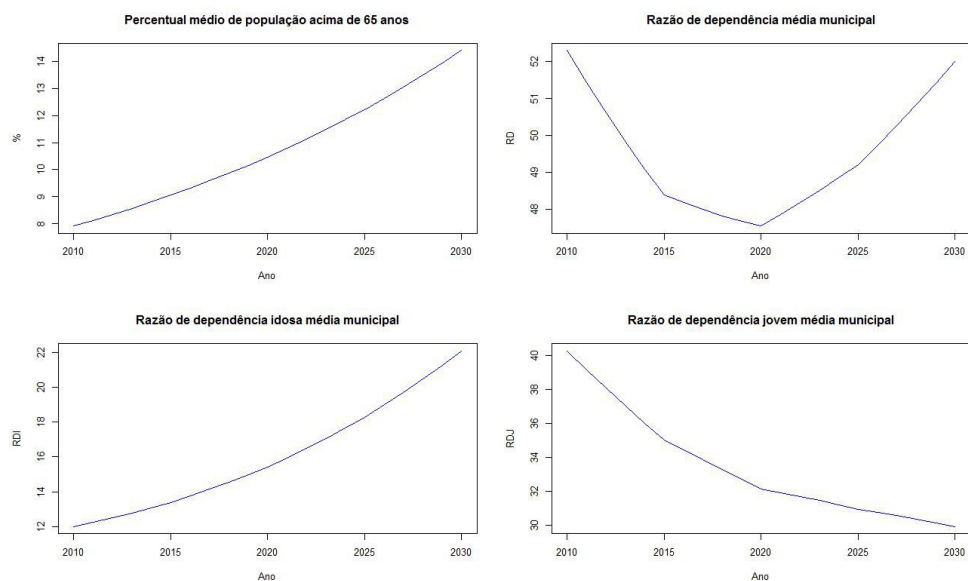
Fonte: Projeções municipais elaboradas a partir de dados básicos dos Censos demográficos 2000 e 2010, e projeções populacionais no nível estadual, revisão 2018, IBGE.

Figura 4
Tendência temporal da média municipal de alguns indicadores populacionais, no período 2010 – 2030. População feminina, Brasil



Fonte: Projeções municipais elaboradas a partir de dados básicos dos Censos demográficos 2000 e 2010, e projeções populacionais no nível estadual, revisão 2018, IBGE.

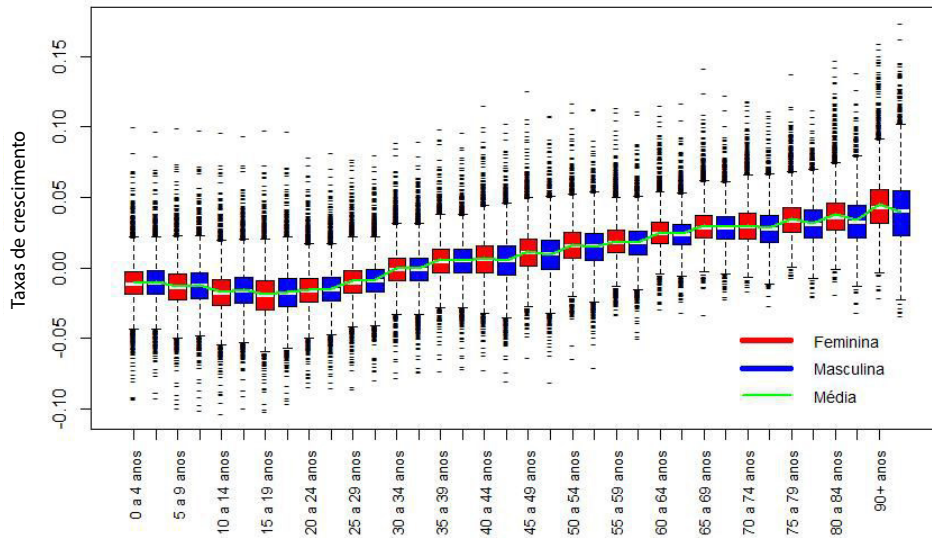
Figura 5
Tendência temporal da média municipal de alguns indicadores populacionais, no período 2010 – 2030. População masculina, Brasil



Fonte: Projeções municipais elaboradas a partir de dados básicos dos Censos demográficos 2000 e 2010, e projeções populacionais no nível estadual, revisão 2018, IBGE.

Uma consequência importante do envelhecimento da estrutura etária é uma redução nas taxas de crescimento populacional. A Figura 6 mostra a distribuição das taxas médias anuais de crescimento exponencial por grupos de idade nos 5.570 municípios brasileiros entre 2010 e 2030. Essa figura reforça o processo de envelhecimento populacional generalizado nas áreas menores do Brasil. Metade dos municípios brasileiros apresentará taxa de crescimento populacional negativa para população abaixo dos 35 anos até 2030, tanto para as mulheres quanto para os homens. Por outro lado, chama a atenção o fato de haver município com taxas de crescimento negativas em grupos etários avançados.

Figura 6
Boxplot das taxas de crescimento entre 2010 e 2030 dos municípios do Brasil por grupos etários e por sexo



Fonte: IBGE (1991, 2000, 2010).

Conforme mencionado anteriormente, outros gráficos com resultados mais detalhados para cada estado do Brasil, assim como as rotinas em R para replicação dos resultados e aplicação do método aqui proposto em outras populações podem ser consultados no site de Demografia UFRN (s. d.). No site disponibilizamos as rotinas em R e acesso ao *GitHub* para usuários interessados em explorar a técnica ou fazer adaptações para aplicações em outras populações.

Considerações finais

Neste artigo propomos uma suavização por um estimador bayesiano empírico nos fatores diferenciais de crescimento do método de relação de coortes para projeções populacionais por sexo e grupos de idade em áreas menores. Este estimador incorpora distribuição espacial de áreas vizinhas e reduz flutuações excessivas no cálculo desses fatores.

Estudos utilizando a abordagem bayesiana para estimar taxas, razões e proporções são relativamente recentes em demografia (Lee, 1997). Uma tentativa de resolver as limitações nos cálculos dos fatores diferenciais de crescimento do método de relação de coortes foi desenvolvida por Assunção (2002). Com base numa metodologia bayesiana hierárquica o autor propôs um modelo estocástico para mensurar a variabilidade dos fatores diferenciais de crescimento em áreas menores. Como resultado, o estudo apontou certa superioridade do modelo bayesiano comparado com outras aplicações do método de relação de coortes para projeções municipais no estado do Paraná (Ipardes e IBGE, 1999). Após decompor a variabilidade dos fatores diferenciais de crescimento em duas componentes, grupos de idade e área, Assunção (2002) concluiu que o efeito de área é maior que o efeito de idade. Ou seja, como antecipamos, a incorporação da distribuição espacial de áreas vizinhas no cálculo dos fatores diferenciais de crescimento do método de relação de coortes é preponderante para projeções populacionais em áreas menores.

A aplicação do método para os municípios do Brasil entre 2000 e 2010 permitiu-nos concluir que a metodologia proposta melhora a precisão das estimativas de população em áreas menores decorrentes das dinâmicas demográficas entre essas áreas. Observou-se uma melhoria na precisão das estimativas tanto para população total quanto para sua distribuição por idade e sexo.

Por outro lado, efeitos produzidos por erros de cobertura ou declaração de idade nos dados de população censitária nas áreas menores não são corrigidos pelo método aqui proposto e podem impactar fortemente as projeções para um logo prazo. Nesses casos, é necessária uma análise e correção nos dados antes de proceder a aplicação do método aqui proposto. A utilização do estimador bayesiano empírico resolve apenas uma das limitações do método de relação das coortes, que é alta variabilidade nas estimativas dos fatores diferenciais de crescimento por grupos de idade e entre áreas vizinhas. O que se fez nesse estudo, no que tange a correção dos dados populacionais, foi comparar a população censitária de 2010 das unidades federativas do Brasil com a população estimada em 2010 pelo IBGE, através do exercício de projeção populacional divulgado por este instituto para as UF, por idade e sexo (IBGE, 2018b).

Aplicações do método para projeções populacionais de todos os municípios do Brasil entre 2010 e 2030 mostraram que o processo de envelhecimento populacional ocorre de forma generalizada no Brasil, com 75 % dos municípios apresentando previsão de taxa de crescimento populacional negativa para a população abaixo dos 25 anos, conforme resultado apresentado na Figura 6. Por outro lado, essa mesma figura apresenta municípios com taxas anuais médias de crescimento populacional negativas para a população dos dois últimos grupos etários, 85 a 89 e 90 anos e mais. Este resultado é intrigante, e deve ser relativizado, uma vez que são estratos populacionais com população ainda menores, portanto mais sujeitos às flutuações aleatórias, ainda que o método de estimação utilizado, como estimador bayesiano empírico, tenha obtido maiores ganhos, em termos de diminuição nos erros de projeção, justamente nesses grupos etários finais.

Por fim, ressaltamos que a combinação de técnicas de estatística bayesiana com métodos demográficos para projeções populacionais é promissora e deve ser considerada pelos demógrafos quando o objetivo é produzir projeções populacionais por sexo e idade em áreas menores. Porém, ressaltamos a necessidade de aplicações e aperfeiçoamentos futuros do método proposto buscando melhorar a precisão nas estimativas produzidas pelo método demográfico de projeções populacionais em áreas menores.

Agradecimentos

Este trabalho contou com o financiamento e apoio administrativo da Secretaria Especial de Assuntos Estratégicos (SAE) da Presidência da República e do Programa das Nações Unidas para o Desenvolvimento do Brasil (PNUD-Brasil), tendo sido implementado pelo Centro Internacional de Políticas para o Crescimento Inclusivo (IPC-IG). Os autores Flávio H. M. A. Freire e Marcos R. Gonzaga agradecem o suporte financeiro do CNPq (Bolsas de Produtividade em Pesquisa – PQ, processos 312609/2018-3 e 307467/2018-0). Marcos R. Gonzaga agradece ao suporte de pesquisa do CNPq (Edital Universal 14/2013).

Referências

- Alves, M. (2016). *Projeções municipais no Brasil com o método de relação de coortes: uma avaliação de alternativas para estimar o fator K*. Monografia de graduação. Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil.
- Araújo, J. D. (2012). Polarização epidemiológica no Brasil. *Epidemiologia e Serviços de Saúde*, 21(4), 533-538. doi: 10.5123/S1679-49742012000400002
- Arriaga, E. (2001). *El análisis de la población con microcomputadoras*. Tese de doutorado. Universidad Nacional de Córdoba, Córdoba, Argentina.
- Assunção, R.M (2002). Método bayesiano de relação de coortes para projeções de pequenas áreas. *Anais do Encontro da Associação Brasileira de Estudos Populacionais*, (8), Ouro Preto, MG, Brasil.
- Barros, L. F. W. e Cavenaghi S. M. (2017). Potencialidades e desafios no uso de registros administrativos nas estimativas populacionais municipais: uma análise exploratória para o Brasil no período 1990-2010. *Anais do Encontro Nacional de Estudos Populacionais*, (20), Foz do Iguaçu, PR, Brasil.
- Brito, L. P. G., Cavenagui S. M. e Jannuzzi P.M. (2010). Estimativas e projeções populacionais para pequenos domínios: uma avaliação da precisão para municípios do Rio de Janeiro em 2000 e 2007. *Revista Brasileira de Estudos de População*, 27(1), 35-57. doi: 10.1590/S0102-30982010000100004
- Castro, M. C. e Simões, C. (2009). Spatio-temporal trends of infant mortality in Brazil. *Anais IUSSP International Population Conference*, (26), Marrakech, Morocco.
- Cavenagui, S. (2012). *Estimaciones y proyecciones de población en América Latina: desafíos de una agenda pendiente*. Rio de Janeiro, Brasil: ALAP.
- Corrêa, C. S., Myrrha, L. J. D. e Fígoli, M. G. B. (2011). Métodos AiBi e logístico para projeção de pequenas áreas: uma aplicação para microrregião de Angicos – RN. *International Seminar on Population Estimates and Projections: Methodologies, Innovations and Estimation of Target Population applied to Public Policies*. Rio de Janeiro, CIC, IBGE. Recuperado de http://www.alapop.org/alap/Docs/ProjectionsSeminar/FinalPresentations/Presentation_RioNov2011_Correa&co-authors.pdf
- Cunha, J. M. P. e Baeninger, R. (2005). Cenários da migração no Brasil dos anos 90. *Caderno CRH - Centro de Recursos Humanos da UFBA*, 18(43), 87-101. doi: 10.9771/ccrh.v18i43.18513
- Demografia UFRN (s. d.). Programa de Pós-Graduação em Demografia – UFRN. Recuperado de <https://demografiufrn.net/laboratorios/lepp>

- Duchesne, L. (1989). Proyecciones de población por sexo y edad para áreas intermedias y menores: método "relación de cohortes". En M. P. Granados. (Ed.), *Métodos para proyecciones subnacionales de población* (pp. 71-126). Bogotá: Celade. Recuperado de https://repositorio.cepal.org/bitstream/handle/11362/9190/S8900623_es.pdf?sequence=1
- González, L. M. e Torres, E. (2012). Estimaciones de población en áreas menores en América Latina: revisión de métodos utilizados. Em S. Cavenaghi. (Org.), *Estimaciones y proyecciones de población en América Latina: desafíos de una agenda pendiente* (pp.105-137). Serie e investigaciones, (2). Rio de Janeiro: ALAP. doi: 10.13140/2.1.2341.3121
- Granados, M. P. (Org.). (1989). *Métodos para proyecciones subnacionales de población*. Bogotá: Cepal, Celade. Recuperado de https://repositorio.cepal.org/bitstream/handle/11362/9190/S8900623_es.pdf?sequence=1
- Hamilton, C. H. e Perry, J. (1962). A short method for projecting population by age from one decennial census to another. *Social Forces*, 41(2), 163-170. doi: 10.2307/2573607
- IGBE (Instituto Brasileiro de Geografia e Estatística)(2018a). *Censo Demográfico 2010 Resultados do Universo*. Rio de Janeiro, Brasil: IGBE. Recuperado de <https://sidra.ibge.gov.br/pesquisa/censo-demografico/demografico-2010/inicial>
- IGBE (Instituto Brasileiro de Geografia e Estatística) (2018b). Projeções da População: Brasil e Unidades da Federação. Revisão 2018. *Série Relatórios Metodológicos*, (40). Rio de Janeiro: IGBE.
- Ipardes e IBGE (Instituto Paranaense de Desenvolvimento Econômico e Social, e Instituto Brasileiro de Geografia e Estatística) (1999). *Projeções de população por sexo e idade 1991-2020*. Curitiba/Rio de Janeiro, Brasil: Ipardes, IBGE.
- Jannuzzi, P. M. (2006). *Projeções populacionais para pequenas áreas: método e aplicações*. Rio de Janeiro, Brasil: Escola Nacional de Ciências Estatísticas.
- Jardim, M. de L. T. (2000). Metodologias de estimativas e projeções populacionais para áreas menores: a experiência do Rio Grande do Sul. *Anais do Encontro Nacional de Estudos Populacionais*, (12), Caxambu, MG, Brasil.
- Justino, J. R., Freire, F. H. M. A. e Lucio, P. S. (2012). Estimação de sub-registros de óbitos em pequenas áreas com os métodos bayesiano empírico e algoritmo EM. *Revista Brasileira de Estudos de População*, 29(1), 87-100. doi: 10.1590/S0102-30982012000100006
- Lee Es, K. e Eldridgeh, T. (1957). *Population redistribution and economic growth, United States, 1870-1950*. Philadelphia, USA: The American Philosophical Society.
- Lee, P. M. (1997). *Bayesian statistics: An introduction*. New York: John Wiley & Sons Inc.
- Madeira, J. L. e Simões, C. S. (1972). Estimativas preliminares da população urbana e rural segundo as Unidades da Federação, de 1960/1980 por uma nova metodologia. *Revista Brasileira de Estatística*, 33(129), 3-11.
- Marshall, R. (1991). Mapping disease and mortality rates using empirical Bayes estimators. *Applied Statistics*, 40(2), 283-294. doi: 10.2307/2347593
- Moura, H. A. (1980). O balanço migratório do Nordeste no período 1950/70. Em H. A. Moura. (Coord.), *Migração interna, textos selecionados*. Fortaleza: BNB/ETENE.
- Naciones Unidas (1975). Métodos para hacer proyecciones de la población urbana y rural. Manual VIII. *Estudios de Población*, (55). New York: Departamento de Asuntos Económicos y Sociales.

- Oliveira, L. A. P. e Oliveira, A. (Orgs.). (2011). *Reflexões sobre os deslocamentos populacionais no Brasil*. Série Estudos e Análises: Informação Demográfica e Socioeconômica, (1). Rio de Janeiro: IBGE.
- Pacheco, C. A. e Patarra, N. (1997). Movimentos migratórios anos 80: novos padrões? Em N. Patarra. (Org.), *Migrações, condições de vida e dinâmica urbana. São Paulo 1980-1993*. Campinas: UNICAMP/IE.
- Paes, N. A. e Albuquerque, M. (1999). Avaliação da qualidade dos dados populacionais e cobertura dos registros de óbitos para as regiões brasileiras. *Revista de Saúde Pública*, 33(1), 33-43. doi: 10.1590/S0034-89101999000100006
- Potter, J. E., Schmertmann, C., Assunção, R. e Cavenagui, S. (2010). Mapping the timing, pace, and scale of the fertility transition in Brazil. *Population and Development Review*, 36(2), 283-307. doi: 10.1111/j.1728-4457.2010.00330.x
- Prata, P. R. (1992). A transição epidemiológica no Brasil. *Cadernos de Saúde Pública*, 8(2), 168-175. doi: 10.1590/S0102-311X1992000200008
- Rao, J. N. K. (2003). *Small area estimation*. Hoboken, USA: Wiley.
- Rees, P. (1989). *Population projections for development planning*. Relatório School of Geography, Leeds University. Recuperado de <http://hdl.handle.net/10068/498978>
- Santos, R. O. e Barbieri, A. (2015). Projeções populacionais em pequenas áreas: uma avaliação comparativa de técnicas de extrapolação matemática. *Revista Brasileira de Estudos de População*, 32(1), 139-163. doi: 10.1590/S0102-30982015000000008
- Schramm, J. M. A., Oliveira, A., Leite, I., Valente, J., Gadelha, A., Portela, M. e Campos, M. (2004). Transição epidemiológica e o estudo de carga de doença no Brasil. *Ciência & Saúde Coletiva*, 9(4), 897-908. doi: 10.1590/S1413-81232004000400011
- Smith, S., Tayman, J. e Swanson, D. A. (2001). *State and local population projections: Methodology and analysis*. New York: Kluwer.
- Sprague, W. W. (2013). *Wood's method: Method for fitting Leslie matrices from age-sex population data, with some practical application*. Dissertation. University of California, Berkeley.
- Swanson, D. A. e Tedrow, L. (2013). Exploring stable population concepts from the perspective of cohort change ratios. *The Open Demography Journal*, 6, 1-17. doi: 10.2174/1874918601306010001
- Waldvogel, B. C. (1998). *Técnicas de projeção populacional para planejamento regional*. Belo Horizonte: UFMG/ CEPDEPLAR.
- Wong, L. L. R. e Carvalho, J. A. (2006). O rápido processo de envelhecimento populacional do Brasil: sérios desafios para as políticas públicas. *Revista Brasileira de Estudos de População*, 23(1), 5-26. doi: 10.1590/S0102-30982006000100002

Apêndice

Demonstração para $E(K_{i,1})$

$$E(K_{i,1}) = E[E(K_{i,1}|\theta_i)] = E\left[E\left(\frac{{}_5P_{x+5,i}^{t+10}}{ESP_i} \middle| \theta_i\right)\right] = E\left[\frac{1}{ESP_i} E({}_5P_{x+5,i}^{t+10}|\theta_i)\right]$$

Da equação 19, temos que:

$${}_5P_{x+5,i}^{t+10} \sim \text{poisson}(ESP_{i,1} \cdot \theta_i) \text{ e que } E({}_5P_{x+5,i}^{t+10}|\theta_i) = ESP_{i,1} \cdot \theta_i.$$

Logo,

$$E(K_{i,1}) = E \left[\frac{1}{ESP_i} ESP_{i,1} \cdot \theta_i \right] = E[\theta_i]$$

$E(K_{i,1}) = m$ De Marshall (1991), θ_i tem distribuição *a priori* não especificada, $E(\theta_i) = m$ e $v(\theta_i) = A$.

Demonstração para $v(K_{i,1})$

$$v(K_{i,1}) = v[E(K_{i,1}|\theta_i)] + E[v(K_{i,1}|\theta_i)]$$

Da demonstração anterior foi visto que: $E(K_{i,1}|\theta_i) = \theta_i$

Então,

$$v(K_{i,1}) = v(\theta_i) + E[v(K_{i,1}|\theta_i)]$$

$$v(K_{i,1}) = v(\theta_i) + E \left[v \left(\frac{{}_5P_{x+5,i}^{t+10}}{ESP_i} \middle| \theta_i \right) \right] = v(\theta_i) + E \left[\frac{1}{(ESP_i)^2} v({}_5P_{x+5,i}^{t+10} | \theta_i) \right]$$

Numa distribuição de Poisson, $E({}_5P_{x+5,i}^{t+10} | \theta_i) = v({}_5P_{x+5,i}^{t+10} | \theta_i)$

Novamente, da equação 19, temos que:

$${}_5P_{x+5,i}^{t+10} \sim \text{poisson}(ESP_{i,1} \cdot \theta_i) \text{ e } E({}_5P_{x+5,i}^{t+10} | \theta_i) = ESP_{i,1} \cdot \theta_i \text{ e}$$

Retomando,

$$v(K_{i,1}) = v(\theta_i) + E \left[\frac{1}{(ESP_i)^2} v({}_5P_{x+5,i}^{t+10} | \theta_i) \right] = v(\theta_i) + E \left[\frac{1}{(ESP_i)^2} ESP_{i,1} \cdot \theta_i \right]$$

$$v(K_{i,1}) = v(\theta_i) + \frac{1}{ESP_{i,1}} E(\theta_i)$$

$$\boxed{v(K_{i,1}) = A + \frac{m}{ESP_{i,1}}}$$

Demonstração para A

Falta conhecer os estimadores das variâncias de θ_i e $K_{i,1}$ o que equivale a conhecer \hat{A} . Tomando s^2 como o estimador para a variância do fator K ponderada pelo tamanho da área, temos:

$$s^2 = \frac{\sum_{i=1}^n {}_5P_{x-5,i}^t (K_{i,1} - m)^2}{T},$$

onde $T = (\sum_{i=1}^n {}_5P_{x-5,i}^t)$ e n é a quantidade de pequenas áreas numa determinada UF.

Vamos tomar duas expressões:

$$E(K_{i,1} - m)^2 = V(K_{i,1}) = A + \frac{m}{ESP_{i,1}} \quad (I)$$

$$E(s^2) = \frac{1}{T} \sum_{i=1}^n {}_5P_{x-5,i}^t E(K_{i,1} - m)^2 \quad (II)$$

Agora, substituindo (I) em (II), temos:

$$E(s^2) = \frac{1}{T} \sum_{i=1}^n {}_5P_{x-5,i}^t \left(A + \frac{m}{ESP_{i,1}} \right) = \frac{1}{T} \sum_{i=1}^n \left({}_5P_{x-5,i}^t A + {}_5P_{x-5,i}^t \frac{m}{ESP_{i,1}} \right)$$

$$E(s^2) = \frac{A}{T} \sum_{i=1}^n {}_5P_{x-5,i}^t + \frac{m}{T} \sum_{i=1}^n \frac{{}_5P_{x-5,i}^t}{ESP_{i,1}}$$

Como já foi dito, $T = (\sum_{i=1}^n {}_5P_{x-5,i}^t)$. Portanto,

$$E(s^2) = A + m \sum_{i=1}^n \frac{{}_5P_{x-5,i}^t / T}{ESP_{i,1}}$$

A partir da expressão acima, estimamos A por:

$$\hat{A} = s^2 - \hat{m} \left[\sum_{i=1}^n \frac{{}_5P_{x-5,i}^t / T}{ESP_{i,1}} \right]$$

Como fica o estimador bayesiano empírico θ_i para o $K_{i,2}$ na fórmula

$${}_5P_{x+10,i}^{t+10} \sim \text{poisson}(ESP_{i,2} \cdot \theta_i)$$

onde θ_i é o risco da população da coorte ser maior ou menor que o esperado sob a hipótese de que a área menor i possua a mesma razão de sobrevivência que a área maior. Com isso, tem-se que a média condicional de ${}_5P_{x+10,i}^{t+10}$ é $E({}_5P_{x+10,i}^{t+10} / \theta_i) = ESP_{i,2} \theta_i$. O estimador de máxima-verossimilhança para θ_i é o fator $K_{i,2}$.

O estimador bayesiano empírico para $K_{i,2}$ é dado por:

$$\hat{\theta}_i = \hat{m} + \hat{c}_i (K_{i,2} - \hat{m})$$

onde,

$$\hat{m} = \frac{\sum_{i=1}^n {}_5P_{x+10,i}^{t+10}}{\sum_{i=1}^n ESP_{i,2}}$$

n indica o número de pequenas áreas e \hat{m} deve ser interpretado como sendo o fator K médio para todo o conjunto de áreas, ou seja, a área maior. Portanto, \hat{m} é 1. O fator de contração:

$$\hat{A} = s^2 - \hat{m} \left[\sum_{i=1}^n \frac{{}_5P_{x,i}^t / T}{ESP_{i,2}} \right]$$

$$\hat{c}_i = \frac{v(\theta_i)}{v(K_{i,2})} = \frac{\hat{A}}{\hat{A} + \frac{\hat{m}}{ESP_{i,2}}}$$

onde,

$$s^2 = \frac{\sum_{i=1}^n {}_5P_{x,i}^t (K_{i,2} - \hat{m})^2}{T}$$

$$T = (\sum_{i=1}^n {}_5P_{x,i}^t)$$