

PROJEÇÕES E CENÁRIOS DEMOGRÁFICOS

21.1 INTRODUÇÃO

Determinar o tamanho e a composição de uma população no futuro ou mesmo no presente (em situações onde uma enumeração direta não é factível), seja por objetivos econômicos, políticos, sociais ou ambientais, é de grande importância para o planejamento e o delineamento de ações estratégicas, tanto no setor público como no privado. Projeções populacionais podem ser utilizadas, por exemplo, para prever a necessidade de serviços de escolas e hospitais, para prever os recursos necessários para o pagamento de benefícios previdenciários ou da folha de pessoal de uma empresa, ou ainda para prever o mercado para determinados produtos (Martine e Lima, 1987).

Por vezes, o objetivo da projeção é simplesmente analisar os determinantes de mudanças na composição futura da população mostrando, em alguns casos, o efeito pela combinação de diferentes alternativas (Smith, Tayman e Swanson, 2001). Isso é conhecido como a construção de *cenários*, ou seja, trajetórias hipotéticas decorrentes da manipulação de um ou outro suposto. A pergunta básica de um cenário é: “O que aconteceria se ?” Isso pode ser útil para a análise dos efeitos da adoção de certa política pública, por exemplo, antes da sua implementação.

Entretanto, na maioria das situações o objetivo da projeção é produzir previsões realistas e baseadas na situação atual, para dar suporte ao planejamento de políticas públicas ou melhorar a tomada de decisões para investimentos futuros. Há ainda situações em que as projeções de população são utilizadas como insumos para a projeção de outras variáveis como, por exemplo, matrículas escolares, os consumidores de energia elétrica ou a demanda por moradias e infraestrutura. Estas projeções são chamadas *funcionais*. A iniciativa privada usa as projeções demográficas, junto com

outros dados, para prever a demanda pelos seus produtos, sejam eles condomínios de luxo, centros comerciais ou produtos de consumo cotidianos cuja demanda depende fortemente da composição demográfica por idade ou localização geográfica.

Embora sejam parecidas, há uma diferença entre *previsões* (ou *prognósticos*), *projeções* e *estimativas*. Uma previsão é simplesmente uma afirmação sobre o futuro. Ela não precisa basear-se numa metodologia demográfica; pode ser simplesmente um “chute” baseado em algum pressentimento intuitivo. Já uma projeção pressupõe uma metodologia explícita para chegar a uma previsão, baseada em algum tipo de dado empírico, embora não necessariamente um dado demográfico. Exemplos disso são as projeções baseadas em dados socioeconômicos sobre as tendências da demanda por certos tipos de bens e serviços, as chamadas projeções por *métodos estruturais*. Uma *projeção* é considerada *demográfica* quando se baseia em dados demográficos de algum tipo, como as funções de mortalidade e fecundidade ou os dados sobre o tamanho da população em censos demográficos sucessivos.

Finalmente, uma projeção é diferente de uma *estimativa* na medida em que uma estimativa não se refere necessariamente ao futuro. Geralmente o termo “estimativa” é reservado para situações onde o valor estimado poderia, em princípio, ser observado diretamente, mas onde a falta de dados suficientemente detalhados ou atualizados faz com que seja necessário recorrer a algumas suposições. Seguindo Smith, Tayman e Swanson (2001), *estimativa populacional* se refere ao exercício de se determinar o tamanho ou composição da população intercensitária. A metodologia para obter estimativas pode envolver o uso de projeções. Por exemplo, a suposição pode ser que o crescimento dos últimos 3 anos, depois do último censo, tenha sido igual ao crescimento médio dos 10 anos anteriores, entre os censos, e uma projeção pode ser feita com base nisso para estimar a população atual.

Já as técnicas empregadas nas projeções podem ser por modelos *subjetivos* ou *objetivos*. Os métodos objetivos especificam algoritmos explícitos que outros pesquisadores podem reproduzir ou modificar. Nos métodos subjetivos, os algoritmos não são explicitamente definidos. Um exemplo são os métodos do tipo Delphi, onde um grupo de experts emite opiniões sobre a trajetória esperada de alguma variável (geralmente não a população, mas a fecundidade ou a migração) que vão sendo ajustadas num processo coordenado de diálogo. Os modelos objetivos se classificam em três grupos: métodos de *interpolação ou extrapolação* (envolvendo modelos matemáticos/estatísticos), modelos *estruturais* (Smith, Tayman e Swanson, 2001; Rao (2003) e Alho, 2005) e *coortes componentes*. Os modelos estruturais, que não serão discutidos aqui, dependem da relação observada entre as variáveis demográficas e outros tipos de dados (uso do solo, emprego etc.) e projetam a população com base na evolução dessas outras variáveis, tipicamente por meio da análise de regressão ou outras técnicas estatísticas semelhantes. Um exemplo deste tipo de procedimento estrutural pode ser encontrado em Santos (2010) e Barbieri e Santos (2011), que projetaram uma região de Minas Gerais a partir de indicadores de emprego. Embora existam situações em que essas outras variáveis sejam mais previsíveis do que a população subjacente, o oposto é mais comum, ou seja, a população é projetada primeiro e as outras variáveis são derivadas a partir daí.

Os métodos de interpolação ou extrapolação baseiam-se na continuação de tendências observadas historicamente e variam desde modelos matemáticos mais simples, como os modelos de crescimento linear, geométrico ou exponencial, até modelos estatísticos mais complexos, como o

modelo logístico ou de regressão por séries temporais. Em contrapartida, os modelos estruturais baseiam-se nas relações entre variáveis demográficas e não demográficas. Nesses modelos, o crescimento populacional (ou alguma de suas componentes) é relacionado com uma ou mais variáveis independentes (geralmente econômicas). Já o método das coortes componentes agrega projeções separadas das 3 componentes demográficas – fecundidade, mortalidade e migração – na projeção da população total por idade e sexo.

Dentre os fatores que definem a escolha deste ou daquele método de projeção, destacam-se os níveis de desagregação regional, de composição populacional e dos objetivos a serem alcançados com a projeção. Os métodos de extrapolação requerem poucos dados e possuem maior aplicabilidade quando há interesse em projetar apenas o volume total da população, geralmente de uma área menor em termos de tamanho da população (como municípios ou distritos). Quando o objetivo é projetar a população por sexo e grupos de idade, o método das coortes componentes é o mais indicado, pelas razões já descritas anteriormente. Entretanto, sua aplicabilidade é limitada à disponibilidade e qualidade dos dados censitários ou de inquéritos e registros vitais.

Existem várias metodologias demográficas especificamente voltadas a pequenas populações (ou populações de pequenas áreas) onde a aplicação das metodologias tipicamente aplicadas a populações maiores pode ser problemática. As mais comuns são as utilizadas para a estimativa do tamanho populacional com base nas informações de populações de áreas maiores que contenham as pequenas áreas. Os métodos mais conhecidos são o método de (extrapolação da) Distribuição do Crescimento, o Logístico (Corrêa, Myrrha e Fígoli, 2011; IBGE, 2007 b) e o de Relação de Coortes, de Duchesne (1989), que ajustam a população das pequenas áreas à população da área total. Esses métodos serão discutidos sistematicamente na seção 21.7 deste capítulo.

Para o caso onde o objetivo é a estimação da população atual e não a sua projeção para o futuro, existem os métodos de regressão sintomáticos, que utilizam dados de registros administrativos (variáveis sintomáticas) juntamente com dados correlatos do último censo para estimar a população total de uma pequena área (Rao, 2003). Os dados administrativos usados podem ser o número de nascimentos, de mortes, de novas unidades de moradia, de matrículas escolares, entre outras variáveis altamente correlacionadas com o crescimento da população e de suas componentes. Mais especificamente, às vezes se usam as estatísticas locais de nascimentos e óbitos não para fins de projeção, mas como indicadores sintomáticos do tamanho da população (Bogue, 1950). Uma breve introdução a este tema será feita na seção 21.10.

O objetivo deste capítulo é descrever os principais métodos de projeção populacionais tendo como foco algum nível de conhecimento sobre os parâmetros que delinham os estudos de população: o tamanho da população num tempo determinado, sua composição (em especial por sexo e idade), sua distribuição geográfica e as suas mudanças ao longo do tempo. Para tanto, são descritas as metodologias de projeção de cada componente demográfica nessa ordem: fecundidade, mortalidade e migração. Ao longo deste capítulo se dará ênfase ao método das coortes componentes para projeção populacional de um país e/ou de suas grandes áreas (como estados, departamentos ou províncias), o qual se utiliza das projeções das componentes demográficas para projetar a população total. Também serão apresentados os métodos de Distribuição do Crescimento, Logístico e Duchesne, com o objetivo de se projetar a população de pequenas áreas, onde as limitações dos dados e o baixo número de eventos e população exposta dificultam a aplicação do método das coortes componentes.

21.2 MÉTODOS DE PROJEÇÃO OU ESTIMAÇÃO POR EXTRAPOLAÇÃO OU INTERPOLAÇÃO MATEMÁTICA

Os métodos de projeção mais comuns envolvem o uso de interpolações ou extrapolações, seja diretamente, para projetar a população, ou indiretamente, para projetar as componentes. A interpolação/extrapolação mais simples é a linear. Por exemplo, se no ano t a migração líquida foi de $M(t)$ indivíduos e no ano $t+1$ foi $M(t+1)$, poder-se-ia supor que no ano $t+2$ será $2 M(t+1) - M(t)$ e no ano $t+3$ será $3 M(t+1) - 2 M(t)$. Às vezes a população total também é projetada de forma linear, como $P(t_p) = P(t+1) + (t_p - t - 1) (P(t+1) - P(t))$, onde t_p representa o momento da projeção.

A limitação principal da extrapolação linear é que ela pode levar a resultados negativos, o que para a maioria dos indicadores demográficos (a migração líquida é uma exceção) não é aceitável. As interpolações/extrapolações mais usadas na prática são a exponencial, a logística e a de Gompertz, que evitam esse problema. A forma mais fácil de aplicar essas interpolações ou extrapolações é por meio dos seguintes passos:

1. No caso da função logística ou de Gompertz, que por definição variam entre 0 e 1, normalizar a variável original pela subtração do limite mínimo I e a divisão pela diferença $S - I$ do limite máximo S e o mínimo I , para garantir que a variável a ser projetada, $y(t)$, varie só entre 0 e 1. I e S não são valores observados, mas limites teóricos que delimitam os valores possíveis que a função pode assumir. Por exemplo, ao projetar a TFT, pode ser razoável – em certos contextos – supor que esta nunca será menor de 0,6 ou maior de 4, de modo que $y(t) = (TFT(t) - 0,6) / (4 - 0,6)$.
2. Linearizar a relação por meio de uma transformação apropriada. Por exemplo, se $y(t)$ tem um comportamento exponencial, $\ln(y(t))$ é uma função linear do tempo. A transformação que lineariza um perfil logístico é o *logito*:

$$\text{logito}(y(t)) = \frac{1}{2} \ln(1/y(t) - 1) \quad (21.1)$$

A transformação logito já foi introduzida em (20.12) e (20.13) do Capítulo 20. Da mesma forma, um perfil de Gompertz pode ser linearizado usando a função dupla logarítmica ou *gompito*, já introduzido em (20.30), da seguinte forma:

$$\text{gompito}(y(t)) = -\ln(-\ln(y(t))) \quad (21.2)$$

Para fins das fórmulas abaixo, a transformação linearizadora é notada com o símbolo φ que, dependendo do caso, pode referir-se ao logaritmo, logito ou gompito.

3. Agora que se obteve uma variável linearizada, é possível interpolar ou extrapolar linearmente a partir de dois pontos t e $t+\Delta t$, para obter um terceiro t_p , que indica o tempo da projeção:

$$\varphi(y(t_p)) = \frac{(t+\Delta t - t_p) \varphi(y(t)) + (t_p - t) \varphi(y(t+\Delta t))}{\Delta t} \quad (21.3)$$

Essa fórmula é idêntica à (18.1) do Capítulo 18, com t em vez de x_i , $t+\Delta t$ em vez de x_{i+1} e t_p em vez de x . Se $t < t_p < t+\Delta t$, se trata de *interpolação*; se $t < t+\Delta t < t_p$, o resultado é uma *extrapolação*, mas a fórmula é a mesma em ambos os casos.

4. Aplicar a transformação inversa de φ para converter $y(y(t_p))$ em $y(t_p)$. No caso de um perfil exponencial, a transformação inversa é simplesmente a função exponencial. No caso de um perfil logístico, a função inversa é:

$$\varphi^{-1}(z) = \frac{1}{1 + e^{2z}} \quad (21.4)$$

No caso de um perfil de Gompertz, o equivalente de (21.4) é

$$\varphi^{-1}(z) = \exp(-\exp(-z)) \quad (21.5)$$

O resultado de (21.4) e (21.5), outra vez, é um número entre 0 e 1, que mostra a posição da quantidade projetada em relação ao seu máximo e mínimo potenciais. No caso onde se aplicou uma normalização no passo 1), é preciso fazer uma transformação final, multiplicando por $(S-I)$ e acrescentando I , para voltar à escala original.

Supondo um crescimento exponencial da população, com observações em t e $t+\Delta t$, a fórmula resultante dos passos descritos acima para projetar a população no momento t_p é a seguinte:

$$P(t_p) = P(t) \left(\frac{P(t+\Delta t)}{P(t)} \right)^{\frac{t_p-t}{\Delta t}} = P(t) r^{t_p-t} \quad (21.6)$$

usando a definição da taxa de crescimento instantânea r conforme (7.1.b). A maior limitação do método exponencial de projeção é que, no longo prazo, a fórmula (21.6) cresce além dos limites razoáveis. A fórmula logística evita esse problema ao impor um limite S superior ao crescimento de $P(t)$. Eventualmente pode ser incluído um limite inferior I também ou o limite inferior pode ser considerado igual a zero.

Aplicando o procedimento descrito acima com esses parâmetros e usando $p(t)$ para descrever a versão normalizada de $P(t)$, ou seja $y(t) = (P(t) - I)/(S - I)$, os resultados são os seguintes:

$$\text{logito}(y(t)) = \frac{1}{2} \ln \left(\frac{S-I}{P(t)-I} - 1 \right) \quad (21.7.a)$$

$$\text{logito}(y(t+\Delta t)) = \frac{1}{2} \ln \left(\frac{S-I}{P(t+\Delta t)-I} - 1 \right) \quad (21.7.b)$$

$$\text{logito}(y(t_p)) = \frac{(t+\Delta t - t_p) \text{logito}(y(t)) + (t_p - t) \text{logito}(y(t+\Delta t))}{\Delta t} \quad (21.7.c)$$

$$P(t_p) = I + \frac{S-I}{1 + \exp(2 \text{logito}(p(t_p)))} \quad (21.7.d)$$

Uma estratégia que pode ser mais fácil hoje em dia é escrever $P(t)$ diretamente na forma de uma função logística, com todos os seus parâmetros:

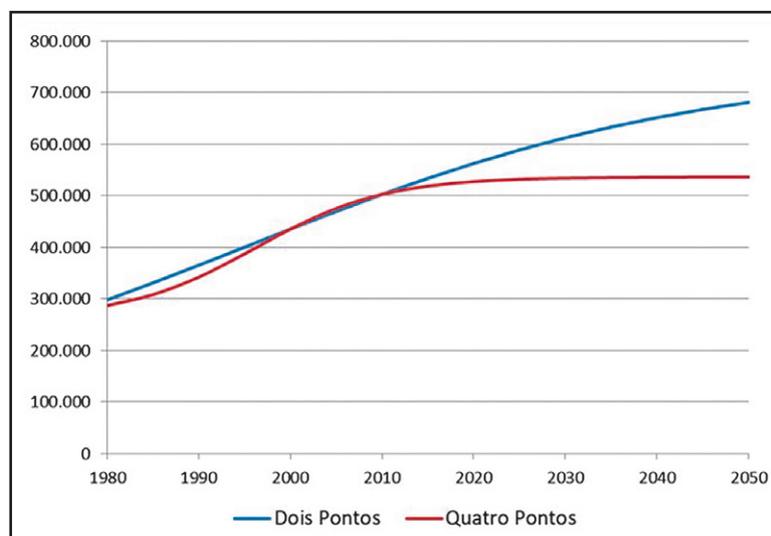
$$P(t) = I + \frac{S - I}{1 + \exp(A + B t)} \quad (21.8)$$

Agora, usando o recurso Solver de EXCEL, escolha A e B de tal maneira que $P(t)$ e $P(t+\Delta t)$ assumam os valores corretos. Por exemplo, Cabo Verde teve uma população de 435.079 em 2000 e 502.382 em 2010. Supondo um limite inferior I de 50.000 e um limite superior S de 750.000, Solver precisa encontrar valores de A e B para os quais

$$\left(50.000 + \frac{700.000}{1 + \exp(A + 2000 B)} - 435.079\right)^2 + \left(50.000 + \frac{700.000}{1 + \exp(A + 2010 B)} - 502.382\right)^2 = 0 \quad (21.9)$$

Aplicando o Solver, se encontra a solução $A = 80,104202$ e $B = -0,04015266$. É possível, inclusive, variar I e S também, de modo que (21.8) precisa ser resolvido para quatro datas (1980, 1990, 2000 e 2010), com uma expressão (21.9) que tem quatro termos. Na medida em que esse procedimento se baseia em mais dados históricos, poder-se-ia supor que a qualidade da extrapolação seria melhor, mas geralmente isso não é o caso. O Gráfico 21.1 mostra as trajetórias interpoladas e extrapoladas de ambas as alternativas. Ambas dão os mesmos resultados em 2000 e 2010, mas a extrapolação de quatro pontos obviamente reproduz corretamente as populações de 1980 e 1990 enquanto a extrapolação de dois pontos as superestima. Entretanto, a trajetória futura implícita na extrapolação de quatro pontos, que sugere um crescimento de quase 0 a partir de 2010, parece implausível.

Gráfico 21.1: Extrapolações logísticas da população de Cabo Verde baseadas em dois pontos (2000 e 2010) e quatro pontos (1980, 1990, 2000 e 2010)



Fonte: Censos de Demográficos Cabo Verde 1980-2010.

21.3 PROJEÇÃO DA FECUNDIDADE

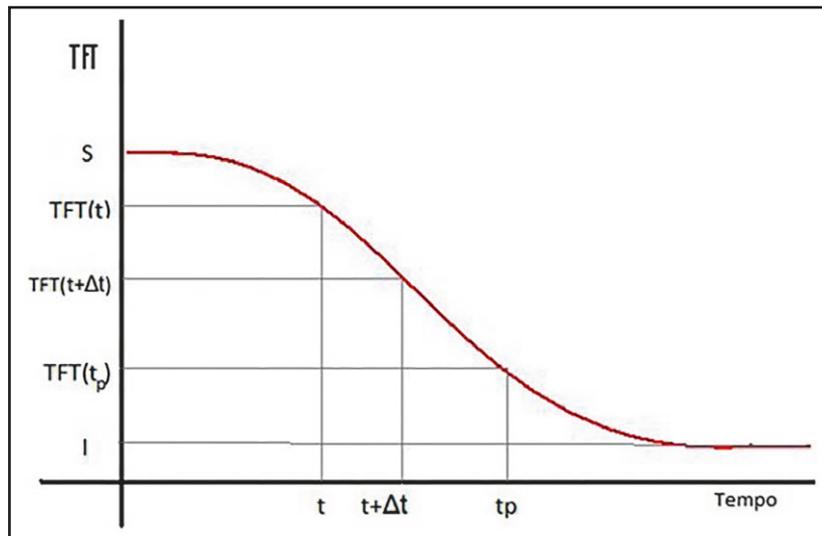
A função logística é usada não só para projetar a população total, mas pode ser usada também para projetar o nível da fecundidade $TFT(t_p)$. O procedimento pode ser aplicado exatamente da mesma forma como em (21.7.a-d) ou da seguinte forma, um pouco mais abreviada

$$TFT(t_p) = I + \frac{S - I}{1 + \exp(a + bt_p)} \quad (21.10.a)$$

$$\text{onde } a = \ln\left(\frac{S - TFT(t)}{TFT(t) - I}\right) \text{ e } b = \frac{1}{\Delta t} \left(\ln\left(\frac{S - TFT(t + \Delta t)}{TFT(t + \Delta t) - I}\right) - a\right) \quad (21.10.b)$$

A Figura 21.1 representa o significado das diferentes quantidades de forma visual.

Figura 21.1: Queda da fecundidade segundo uma função logística



Fonte: Dados fictícios.

O DHS de Timor-Leste fornece um exemplo de como usar a função logística para projetar a TFT. No DHS de 2009-10 ($t=6,5 = 2010,0$), a TFT encontrada foi 5,675. No DHS de 2016 ($t=2016,5$), tinha caído para 4,205. Supondo que a fecundidade máxima S seja igual a 7,0 e a mínima I a 1,8, pode-se, então, calcular os valores de a e b que definirão os níveis de fecundidade a cada tempo.

$$a = \ln\left(\frac{S - TFT(t)}{TFT(t) - I}\right) = \ln\left(\frac{7,0 - 4,205}{4,205 - 1,8}\right) = 0,1503 \quad (21.11.a)$$

$$b = \frac{1}{\Delta t} \left(\ln\left(\frac{S - TFT(t + \Delta t)}{TFT(t + \Delta t) - I}\right) - a\right) = \frac{-1}{6,5} \left(\ln\left(\frac{7,0 - 5,675}{5,675 - 1,8}\right) - 0,1503\right) = 0,1882 \quad (21.11.b)$$

Depois, calcula-se a TFT para cada período, por exemplo 2015-2020:

$$TFT(t_p) = I + \frac{S - I}{1 + \exp(a + bt_p)} = 1,8 + \frac{7,0 - 1,8}{1 + \exp(0,1503 + 0,1882 \cdot 1)} = 3,964 \quad (21.12)$$

Seguindo o mesmo raciocínio para os períodos seguintes,

Ano	2016,5	2017,5	2022,5	2027,5	2032,5	2037,5	2042,5	2047,5
TFT	4,205	3,964	2,932	2,309	2,011	1,884	1,833	1,813

Assim, projetou-se o nível da fecundidade para todos os anos seguintes da projeção. Repare que foram utilizados, como medidas de tempo, o meio do período de cada projeção quinquenal como aproximação para o tempo a que o valor da TFT projetada se refere. O módulo TFRLGST do pacote PASEX pode ser usado para fazer esses ajustes automaticamente.

Para projetar a estrutura da fecundidade parte-se da estimação da função de fecundidade acumulada a cada idade, $F(x)$. Dividindo esta função pela TFT, obtém-se uma quantidade normalizada que varia entre 0 e 1. Além de variar por idade (x), $F(x)/TFT$ também varia no tempo. Para qualquer idade x , $F(x)/TFT$ pode ser projetada no tempo usando um perfil de Gompertz. Portanto:

$$Y(x, t) = -\ln(-\ln(\frac{F(x, t)}{TFT(t)})) \quad (21.13.a)$$

$$Y(x, t + \Delta t) = -\ln(-\ln(\frac{F(x, t + \Delta t)}{TFT(t + \Delta t)})) \quad (21.13.b)$$

$$Y(x, t_p) = \frac{(t + \Delta t - t_p) Y(x, t) + (t_p - t) Y(x, t + \Delta t)}{\Delta t} \quad (21.13.c)$$

$$F(x, t_p) = TFT(t_p) \exp(-\exp(-Y(x, t_p))) \quad (21.13.d)$$

Em princípio, o procedimento (21.13.a-d) pode ser aplicado para qualquer dupla de funções $F(x, t)$ e $F(x, t + \Delta t)$ e o resultado $F(x, t_p)$ sempre será uma função não negativa definida no domínio correto. Entretanto, é preciso ter certo cuidado para evitar que certas irregularidades numa das duas funções sejam amplificadas no momento de fazer a extrapolação. Por isso, é recomendável graduar $F(x, t)$ e $F(x, t + \Delta t)$ antes de fazer a extrapolação. Em princípio isso pode ser feito de diferentes maneiras, mas a mais comum é usando o próprio padrão etário de uma função Gompertz relacional que foi explicado no Capítulo 20. Desta forma, a graduação e a extrapolação podem ser unificadas num único procedimento.

Embora a projeção pela função de Gompertz seja uma solução bastante comum, particularmente na América Latina, é preciso alertar que existem alternativas. Por exemplo, Silva (2002) projetou a fecundidade portuguesa pela extrapolação dos parâmetros de uma distribuição Gama. Em vez de usar diretamente os parâmetros α e β , ela fez a projeção a partir da média $\alpha\beta$ e do desvio-padrão $\alpha^{1/2}\beta$.

O Capítulo 20 especificou dois procedimentos para fazer a graduação de $F(x,t)$ e $F(x,t+\Delta t)$: o método tradicional na Tabela 20.3 e o método baseado no Solver de EXCEL na Tabela 20.4. Aqui se usará o método mostrado na Tabela 20.4, com a diferença de que as estimativas de C que definem as TFT de 2009-2010 e 2016 serão mantidas nos valores anteriormente calculados (5,675 e 4,205), em vez dos valores ligeiramente ajustados que resultam do procedimento da Tabela 20.4 com três parâmetros livres. Ou seja, apenas α e β serão otimizados, mas C , não. Isso é feito puramente por motivos didáticos. Na prática, o procedimento também pode ser aplicado com todos os três parâmetros livres. As taxas específicas a serem digitadas em G2:G8 são as seguintes:

	2009-10	2016
15-19	0,051	0,042
20-24	0,221	0,188
25-29	0,276	0,223
30-34	0,254	0,196
35-39	0,197	0,119
40-44	0,089	0,056
45-49	0,047	0,017
TFT	5,675	4,205

Procedendo dessa forma e com esses dados, se encontram os seguintes valores:

2009-10:	$\alpha = -0,3530$	2016:	$\alpha = -0,2725$
	$\beta = 1,0120$		$\beta = 1,1265$
	$C = 5,675$		$C = 4,205$

Agora a projeção de α e β é uma simples extrapolação linear. Por exemplo, para 2020-24 ($t+6 = 2022,5$), o cálculo é

2020-24:	$\alpha = -0,2725 + 6,0 (-0,2725 - (-0,3530)) / 6,5 = -0,1982$	(21.14.a)
	$\beta = 1,1265 + 6,0 (1,1265 - 1,0120) / 6,5 = 1,2322$	(21.14.b)
	$C = 2,932$	(21.14.c)

Estes valores podem ser substituídos, juntos com o C extrapolado em cima, nas células J1:J3 da planilha de EXCEL mostrada na Tabela 20.4, resultando nas seguintes taxas de fecundidade:

15-19	0,034
20-24	0,146
25-29	0,171

30-34	0,131
35-39	0,077
40-44	0,026
45-49	0,002

As taxas para os outros anos podem ser derivadas da mesma forma.

A forma de aplicação do método descrita acima é como extrapolação, ou seja, especificando a fecundidade em t_p , como resultado da extrapolação do nível e do padrão observado em momentos anteriores t e $t+\Delta t$. Uma forma alternativa é a aplicação do método como uma interpolação entre o momento t e um momento no futuro distante, quando se assume que a fecundidade terá um determinado nível e um determinado padrão etário mais ou menos previsíveis. No caso da fecundidade esta forma de aplicação do método é pouco comum porque é relativamente difícil formular hipóteses plausíveis sobre uma suposta fecundidade limite no futuro distante. Mas no caso da mortalidade, existem mais elementos para poder estipular alguma tábua de vida como o limite para o qual todas as tábuas de vida dentro de uma determinada região geográfica tendem a convergir no longo prazo. Abaixo se descreve como os métodos descritos acima podem ser aplicados à projeção da mortalidade e como podem ser adaptados ao caso da projeção com o uso de uma tábua limite (tema também abordado por Arriaga, 1987).

21.4 PROJEÇÃO DA MORTALIDADE

O mesmo procedimento descrito em (21.13.a-d) também pode ser usado para projetar a mortalidade, com a diferença de que a função projetada é a ℓ_x normalizada (com $\ell_0 = 1$) e que a transformação aplicada geralmente é o logito e não o gompito. Portanto

$$Y(x, t) = \frac{1}{2} \ln \left(\frac{1}{\ell_x(t)} - 1 \right) \quad (21.15.a)$$

$$Y(x, t + \Delta t) = \frac{1}{2} \ln \left(\frac{1}{\ell_x(t + \Delta t)} - 1 \right) \quad (21.15.b)$$

$$Y(x, t_p) = \frac{(t + \Delta t - t_p) Y(x, t) + (t_p - t) Y(x, t + \Delta t)}{\Delta t} \quad (21.15.c)$$

$$\ell_x(t_p) = \frac{1}{1 + \exp(2 Y(x, t_p))} \quad (21.15.d)$$

É de se notar que o procedimento (21.15.a-d) projeta não só o padrão, como no caso da fecundidade, mas também o nível da mortalidade, já que o modelo de logitos tem só 1 parâmetro para descrever o padrão e não 2.

O procedimento usado para a projeção com uma tábua limite é muito parecido com (21.15.a-d). A diferença principal é a substituição de $t+\Delta t$ por t_L , o tempo quando a tábua limite será alcançada, e $\ell_x(t+\Delta t)$ por $\ell_x(t_L)$, a própria tábua limite:

$$Y(x, t) = \frac{1}{2} \ln \left(\frac{1}{\ell_x(t)} - 1 \right) \quad (21.16.a)$$

$$Y(x, t_L) = \frac{1}{2} \ln \left(\frac{1}{\ell_x(t_L)} - 1 \right) \quad (21.16.b)$$

$$Y(x, t_p) = \frac{(t_L - t_p) Y(x, t) + (t_p - t) Y(x, t_L)}{(t_L - t)} \quad (21.16.c)$$

$$\ell_x(t_p) = \frac{1}{1 + \exp(2 Y(x, t_p))} \quad (21.16.d)$$

No contexto dos PALOP, é preciso considerar o efeito da AIDS (SIDA). Por exemplo, Antunes (2008) considera que a projeção demográfica nacional de 1997, quando comparada com os resultados do Censo daquele ano, mostra uma sobre-estimação considerável: 19 milhões, em vez do resultado ajustado de 16.099.246 do Censo. Embora também existam outras razões para a sobre-estimação, Antunes sugere que a epidemia da AIDS (SIDA) teve um papel. Arnaldo e Muamoha (2014) mencionam que, apesar do maior declínio da mortalidade no período 1997-2007 em comparação com o período anterior, as províncias com maior prevalência de HIV (Maputo Cidade, Maputo Província e Gaza) tiveram declínios mais modestos ou até um ligeiro aumento da mortalidade. Especificamente, a cidade de Maputo, devido ao impacto do HIV/SIDA na mortalidade adulta, teve uma redução da sua esperança de vida, de 58,3 anos em 1997 para 56,1 anos em 2007. Os detalhes da projeção da mortalidade por AIDS (SIDA) fogem aos objetivos deste livro, mas vale a pena mencionar que alguns programas de projeção como o DEMPROJ (ver Capítulo 17) já incorporam esta componente.

21.4.1 Projeção pelo Método Lee-Carter

O método Lee-Carter combina um modelo demográfico com um método de séries temporais para extrapolar a mortalidade. O método gera estimativas com uma boa precisão, além de ser de simples implementação para modelar a distribuição das taxas de mortalidade por idade (Lee e Carter, 1992; Lee e Miller, 2001; Booth e Tickle, 2008).

Modela-se as TEMs baseado numa matriz das taxas específicas de mortalidade por idade em escala logarítmica. O primeiro passo do método Lee-Carter consiste em modelar estas taxas como:

$$\ln(M_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t} \quad (21.17)$$

onde a_x , b_x e k_t são parâmetros a serem estimados e $\varepsilon_{x,t}$ é o efeito aleatório.

A solução desta regressão ocorre a partir da decomposição do valor singular (“Singular Value Decomposition” ou SVD, em inglês) no logaritmo da matriz de taxas em cada período da série histórica $M_{x,t}$. O método impõe duas restrições para obter uma solução única:

1. ${}_n a_x$ deve ser calculado a partir da média de $\ln({}_n M_x)$ ao longo do tempo, ou seja, a média do padrão das taxas de mortalidade por idade ao longo da série histórica disponível para a mortalidade;
2. ${}_n b_x$ soma 1 e é a distribuição relativa da taxa de mudança da mortalidade por idade;
3. k_t soma 0 e representa o nível da mortalidade no tempo t .

Os parâmetros ${}_n a_x$ e ${}_n b_x$ são fixados no tempo, enquanto k_t é estimado com o método de séries temporais ARIMA¹. O modelo ARIMA mais apropriado nesse contexto de tendência linear nas taxas específicas de mortalidade por idade é o passeio aleatório (Li et al., 2002; Booth et al., 2004).

O modelo para k_t , portanto, é dado por:

$$k_t = c + k_{t-1} + u_t \quad (21.18)$$

onde c representa a mudança média anual em k_t . Uma vez estimado k_t , é possível derivar todas as taxas específicas de mortalidade por idade para os respectivos anos da projeção.

Com o intuito de tornar o método de Lee e Carter robusto para subpopulações ou localidades com poucos dados, foram realizados desdobramentos do artigo original (Li, Lee e Tuljapurkar, 2007). Verificou-se que o método poderia ser utilizado com pelo menos 2 observações para a projeção da mortalidade e com pelo menos 3 observações para a estimação da variância da mortalidade e, conseqüentemente, do intervalo de confiança para as projeções. Porém, quanto menor a quantidade de pontos no tempo, maior a incerteza e, conseqüentemente, maior será o intervalo de confiança estimado (Li, Lee e Tuljapurkar, 2007). Surgiram outros métodos derivados do método original de Lee e Carter, como apresentado por Shang, Booth e Hyndman (2011) e por Shang (2012).

Vale mencionar que existem algumas variantes do modelo de Lee-Carter. A variante de Renshaw-Haberman (2006) generaliza o modelo Lee-Carter através de uma especificação não linear generalizada para agregar o efeito coorte. Currie introduziu um modelo do tipo idade-período-coorte mais simples, que também é um caso especial do modelo Renshaw-Haberman. Outras alternativas para a especificação de Lee-Carter foram apresentadas por Cairns, Blake e Dowd (2008) e por Plat (2009). Todos esses modelos são discutidos em detalhe por Peixoto, Neves e Melo (2016-17) e por Pereira (2013).

Existe uma extensa literatura sobre o modelo de Lee-Carter e suas aplicações. No Brasil o modelo foi aplicado pela primeira vez por Fígoli (1998). A tese de Santos (2007) mostra outra aplicação e discute o próprio modelo em mais detalhe no Capítulo 3 da sua tese. Silva (2009) usou o modelo para projetar a mortalidade do município de São Paulo de 2006 a 2100, usando dados desde 1920. Moura (2015) fez uma aplicação ao Rio Grande do Sul e Cavalcante e Reinaldo (2017-18) ao Ceará. Silva (2019) aplicou sete variantes do método para grandes regiões do Brasil e produziu séries históricas de taxas de mortalidade que vão de 1950 a 2100. Coelho (2001, 2005 e

¹ Do inglês “autoregressive integrated moving average”, que significa “modelo autorregressivo integrado de média móvel”.

2009) elaborou várias aplicações ao caso de Portugal. Vale mencionar também o programa LCFIT, desenvolvido por W. Webb Sprague, que permite a estimação dos parâmetros do modelo (<http://lcf.it.demog.berkeley.edu/>).

21.5 PROJEÇÃO DA MIGRAÇÃO

A migração é a componente mais imprevisível da dinâmica demográfica. Como foi assinalado no Capítulo 11, é a componente com maior volatilidade, suscetível a variações grandes e repentinas em virtude de mudanças econômicas, políticas, climáticas, entre outras. Arelado a isso, há ainda a dificuldade de dados: não existem séries históricas detalhadas de dados de migração em grande parte dos países. As projeções de migração geralmente se baseiam em supostos extremamente simples ou então têm um caráter mais substantivo. Um suposto frequentemente adotado em projeções nacionais é que a migração internacional manterá o mesmo padrão por sexo e idade, com um nível que tenderá a zero dentro de um prazo a ser especificado. Entretanto, dada a importância crescente da migração internacional no mundo e especificamente em países como o Brasil ou Portugal, esse suposto pode ser cada vez menos realista.

A Revisão de 2019 da Divisão de População das Nações Unidas essencialmente manteve a tendência da migração líquida constante até 2100, exceto num período inicial em que os números líquidos de migrantes são reduzidos para o seu equilíbrio de longo prazo. Tomando como referência as tendências históricas das últimas décadas, esse perfil parece excessivamente conservador. Por outro lado, tomando em conta as reações políticas contrárias à migração estrangeira atualmente observadas em países como os EUA, Itália e Hungria, não é impossível que a tendência nas próximas décadas acabe sendo até mais restritiva. Abel (2018) elaborou cinco cenários alternativos de migração líquida para o mundo e as grandes regiões até 2100, para modelar o impacto de diferentes cenários socioeconômicos e ambientais. Outro modelo recentemente formulado para a projeção da migração internacional é o de Buettner e Muenz (2018), que se baseia na metodologia de múltiplos estados (ver seção 21.4). Atualmente os maiores avanços nesta área se concentram no desenvolvimento de modelos probabilísticos (ver seção 21.7) para estimar e projetar as tendências da migração internacional, com o uso de séries temporais e técnicas bayesianas (por exemplo, Abel et al., 2013; Azose, 2016; Azose e Raftery, 2015).

No Brasil, o IBGE, nas suas projeções publicadas em 2018, fez os saldos migratórios convergirem para um valor constante a partir de um determinado ano para todas as UFs. Em projeções *subnacionais*, em vez de manter os volumes de migração líquida de cada unidade constantes, pode ter mais sentido manter as *taxas* de migração constantes, para considerar o efeito de mudanças nos tamanhos relativos das unidades sobre os volumes de migrantes. Entretanto, fazer isso com as *taxas líquidas* de migração interna tem dois inconvenientes:

1. Como a migração líquida é o resultado de imigração e emigração, a taxa líquida não tem uma interpretação probabilística clara (ver Capítulo 11) de modo que a priori não há razões para supor que ela deveria ser constante.
2. Na medida em que os tamanhos relativos das unidades mudam, taxas constantes de migração líquida levam a volumes cada vez mais inconsistentes, ou seja, volumes cuja soma se afasta cada vez mais do total de 0 que se espera para a migração interna a nível nacional.

Por este motivo, quando se trabalha com taxas constantes, é preferível usar *taxas brutas*. Isso pode ser feito de duas maneiras. A solução mais completa é usar taxas que identificam tanto a origem como o destino. Isso também abre a possibilidade de especificações alternativas em que as taxas variam, por exemplo, em função da dinâmica socioeconômica esperada na origem e no destino. O modelo mais adequado de projeção neste caso é o modelo de múltiplos estados (ver seção 21.6.4). Mas o modelo de múltiplos estados é complexo e difícil de aplicar quando o número de unidades geográficas é grande (por exemplo, as 27 UFs do Brasil). Uma solução mais simples é o uso daquilo que em inglês se chama “migrant pool” (Smith, Tayman e Swanson, 2001: 123) o que em português poderia ser traduzido como *estoque de migrantes* ou *fundo de migrantes* (Irwin, 1987). Nesta metodologia, as taxas de emigração de cada unidade são usadas para calcular o total de emigrantes de todas as regiões conjuntamente, o que se chama o estoque de migrantes. Este estoque depois é dividido entre as unidades para definir a imigração a cada região, de forma proporcional às taxas de imigração (do período base ou projetadas de alguma forma). Procedendo desta forma se obtêm fluxos migratórios consistentes. A única desvantagem é que o método não permite a formulação de hipóteses a interação de regiões específicas, por exemplo a evolução do sistema migratório que liga a região Nordeste com o Estado de São Paulo.

Em teoria, o padrão etário da migração pode ser projetado por meio do modelo de Rogers e Castro (20.37.a-d), mas na prática, o grande número de parâmetros desse modelo dificulta a sua especificação completa para fins de projeção. Além, disso, como o modelo é formulado em termos de taxas brutas de migração ele se presta mais ao modelo de múltiplos estados do que ao modelo tradicional de componentes, baseado na migração líquida.

21.6 PROJEÇÃO DA POPULAÇÃO: O MÉTODO DAS COORTES COMPONENTES

Embora existam outros modelos baseados em extrapolações ou correlações que podem funcionar bem para projetar a população total, a projeção do tamanho de uma população por sexo e idade requer a compreensão de mudanças e tendências nas componentes que regem a dinâmica demográfica ao longo do tempo: fecundidade, mortalidade e migração. Em termos de mudanças na estrutura etária de uma população, reduções continuadas nas taxas de fecundidade fazem com que ocorra uma redução no peso relativo de crianças em comparação com as proporções de adultos e idosos. Assim, a informação do quanto as mulheres estão tendo filhos (ver Capítulos 2 e 9) é de grande relevância para qualquer exercício de projeção populacional. Por sua vez, reduções continuadas na mortalidade (ver Capítulos 2 e 7) também jogam um papel em definir o tamanho e estrutura das populações futuras. Finalmente, segundo a equação de equilíbrio populacional apresentada no Capítulo 7, também é preciso levar em conta o efeito da migração, principalmente a nível local. Os fatores determinantes da mortalidade, fecundidade e migração possibilitam ao demógrafo formular hipóteses e traçar cenários futuros para o comportamento destas componentes demográficas. A seguir se explica como as hipóteses formuladas sobre cada uma das componentes podem ser consolidadas numa projeção da população por sexo e idade e eventualmente outras características, como a residência rural-urbana ou a atividade econômica.

Resumidamente, o Método das Coortes Componentes baseia-se na aplicação da Equação de Equilíbrio Populacional para cada grupo de idade e sexo, ou seja, para cada coorte. Deste modo, o número de pessoas em cada grupo de idade em determinado ano são coortes que serão

acompanhadas ao longo do tempo até o horizonte de projeção. O que determina se uma coorte aumenta ou diminui ao longo do tempo é o comportamento por idade das taxas específicas de fecundidade, mortalidade e migração que atuam em cada intervalo de tempo de projeção. A fortaleza do método reside, assim, no uso de um raciocínio eminentemente demográfico, o que por sua vez, tem como grande vantagem sobre outros, a possibilidade de avaliar a plausibilidade dos cenários compostos pela combinação do comportamento das três variáveis mencionadas. A lógica do método foi formulada há mais de um século (ver, por exemplo, Cannan, 1895), embora os textos mais completos sobre sua formalização sejam, talvez, Leslie (1945) ou Welpton (1936). Existem vários programas de computador (ver seção 17.2 do Capítulo 17) que ajudam a executar projeções por componentes, mas a lógica básica do procedimento pode ser programada facilmente em EXCEL.

Para a realização das projeções, o intervalo de tempo de uma população projetada para a outra deve ser o mesmo utilizado para os grupos etários. Assim, se os grupos etários da população inicial são quinquenais, a projeção só pode ser feita para períodos de 5 em 5 anos; se os grupos etários são de idade simples, a projeção deve ser feita para ano a ano, seguindo os mesmos princípios que definem um diagrama de Lexis, conforme apresentado do Capítulo 7. Neste capítulo, por praticidade, será utilizada a projeção por períodos quinquenais, com a população agrupada em grupos etários quinquenais, em conformidade com os demais capítulos deste livro. Mas também é possível fazer projeções anuais, usando a estrutura de população por idade simples.

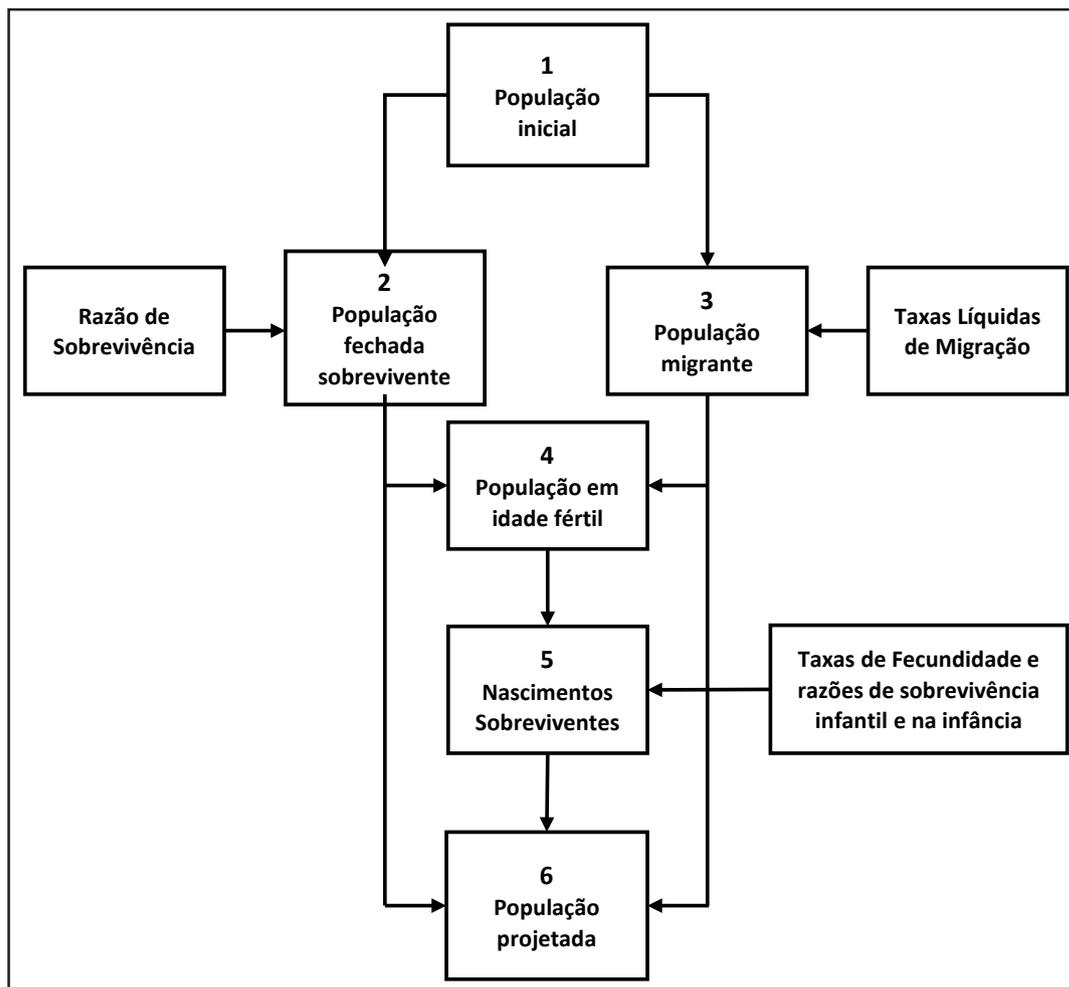
As etapas para implementação do modelo convencional de componentes são:

1. Projeta-se a população de cada subgrupo no início do intervalo etário de forma a estimar o número de sobreviventes no início do intervalo seguinte.
2. Calcula-se o número de nascimentos no intervalo e quantos sobrevivem até o início do intervalo seguinte.
3. Adicionam-se os imigrantes e subtraem-se os emigrantes em cada subgrupo durante o intervalo.
4. Calcula-se o número de nascimentos desses migrantes e seus nascimentos que sobrevivem até o início do intervalo seguinte.

No modelo multirregional, que implementa as componentes de uma forma diferente, a migração não é projetada separadamente, mas integrada diretamente à mortalidade e fecundidade.

Essas etapas são ilustradas no fluxograma da Figura 21.2. Nota-se que a projeção pelo método das coortes componentes requer, para sua implementação, a estimação da estrutura etária das taxas de fecundidade, mortalidade e migração no tempo corrente, além de pressupostos relacionados às mudanças futuras no padrão etário dessas taxas, princípios que norteiam as projeções realizadas por organizações e institutos oficiais de estatísticas de população.

Figura 21.2: Esquema geral de aplicação do método das coortes componentes



Fonte: Adaptado de Smith, Tayman e Swanson (2001).

Seguindo o esquema apresentado na Figura 21.2, a população por grupos quinquenais de idade no ano base deve ser aquela do último censo disponível. Caso existam erros de cobertura ou de declaração de idade no censo demográfico, tais erros precisam ser avaliados e corrigidos antes que se inicie a projeção, conforme apresentado no Capítulo 16.

21.6.1 Efeito da mortalidade

A primeira etapa na aplicação do método é a projeção da população fechada sobrevivente (estágio 2 do esquema apresentado na Figura 21.2). Supondo que a população base foi recenseada no tempo t , a população sobrevivente por grupos quinquenais acima da idade $x=5$ (exceto o último grupo etário aberto) no ano $t+5$ (sem considerar a migração) é obtida por:

$${}_5P_x(t+5) = \frac{{}_5L_x(t, t+5)}{{}_5L_{x-5}(t, t+5)} {}_5P_{x-5}(t) = {}_5S_{x-5}(t, t+5) {}_5P_{x-5}(t) \quad (21.19)$$

onde ${}_5P_x(t+5)$ é a população no grupo etário $(x, x+5)$ no ano $t+5$, ${}_5L_x$ é a função anos-pessoa da tábua de vida projetada para o período $(t, t+5)$ e ${}_5P_{x-5}(t)$ é a população do grupo etário $(x-5, x)$ no ano t . A razão de sobrevivência do grupo etário $(x-5, x)$ para o grupo etário $(x, x+5)$ na tábua de vida se conhece como ${}_5S_{x-5}$.

Para o último grupo etário (aberto), aplica-se a seguinte equação:

$$P_{x+}(t+5) = \frac{{}_5L_x(t, t+5)}{{}_5L_{x-5}(t, t+5)} {}_5P_{x-5}(t) + \frac{T_{x+5}(t, t+5)}{T_x(t, t+5)} P_x(t) \quad (21.20)$$

onde P_{x+} refere-se à população do grupo etário aberto e T_x é a função anos-pessoa vividos pela coorte hipotética a partir da idade x (ver Capítulo 9). Ou seja, irão compor o grupo etário aberto no tempo $t+5$ as pessoas que já estavam no grupo etário aberto no tempo t e que sobreviveram até o tempo $t+5$ mais as pessoas que estavam no grupo etário anterior que sobreviveram até o grupo etário aberto. Para operacionalizar a equação (21.20) é necessário que o último grupo etário aberto da tábua de vida para o período $(t, t+5)$ seja um grupo etário imediatamente acima do grupo etário aberto da população projetada. Caso essa informação não esteja disponível, é preciso reformatar o problema, baixando o limite inferior do intervalo aberto em 5 anos.

21.6.2 Acrescentar a componente migratória

A etapa seguinte consiste em determinar a população migrante (terceiro estágio no esquema da Figura 21.2), tendo por hipótese que as mulheres migrantes, em idade fértil, poderão contribuir (direta ou indiretamente) para o total nascimentos sobreviventes no local de destino. A inclusão dessa população migrante pode ser obtida, aplicando-se à população fechada projetada as taxas líquidas de migração (coluna 5):

$${}_5P_x(t+5) = [{}_5P_x(t) {}_5TLM_x(t, t+5)] + {}_5P_x(t) \quad (21.21)$$

onde ${}_nTLM_x$ são as taxas líquidas de migração por grupos etários e sexo. Este método funciona sem maiores problemas no caso de uma projeção nacional, onde não há necessidade de compatibilizar os fluxos totais de migrantes com a dinâmica demográfica do “resto do mundo”. A situação não é fundamentalmente diferente quando se projeta apenas uma unidade subnacional, como a Província de Gaza em Moçambique ou o Estado de São Paulo no Brasil. Entretanto, quando se projeta o conjunto completo de determinado tipo de unidades dentro de uma população nacional (as UFs do Brasil ou as Províncias de Angola), é preciso distinguir entre a migração interna e externa. Enquanto a migração externa continua não apresentando maiores restrições, a migração interna precisa satisfazer o critério de que a soma da migração interna líquida das unidades precisa ser zero. Isso exige certos cuidados, principalmente em países como o Brasil, onde a migração interna é muito mais volumosa do que a externa. Neste caso, a aplicação da fórmula (21.21) pode levar a resultados pouco plausíveis.

Para ilustrar o problema, suponha-se um país hipotético com três regiões, A, B e C, e suponha que, numa determinada faixa de idade e sexo, as populações iniciais do período de base foram

125.000, 75.000 e 50.000, respectivamente, com TLMs de 10, -10 e -10 por mil, respectivamente. Finalmente, suponha que o crescimento vegetativo anual projetado para as regiões seja 3% para a região A, 1% para a região B e -1% para a região C, constantes no tempo. Nestas condições, a evolução das três populações será a seguinte:

Ano 0 (Ano Base)	125.000 +	75.000 +	50.000 -
Crescimento Vegetativo	3.750 +	750 -	500 -
Crescimento Migratório	1.250 =	750 =	500 =
Ano 1	130.000 +	75.000 +	49.000 -
Crescimento Vegetativo	3.900 +	750 -	490 -
Crescimento Migratório	1.300 =	750 =	490 =
Ano 2	135.200 +	75.000 +	48.020 -
Crescimento Vegetativo	4.156 +	750 -	480 -
Crescimento Migratório	1.352 =	750 =	480 =
Ano 3	135.200 +	75.000 +	47.060 -

Entre o ano 0 e o ano 1 (o período base), o crescimento migratório total das três regiões é $1.250 - 750 - 500 = 0$, ou seja, os fluxos se compensam. Já no segundo período, a soma é $1.300 - 750 - 490 = 60$, ou seja, há uma inconsistência que envolve 60 migrantes. No terceiro período, esta inconsistência aumenta para $1.352 - 750 - 480 = 144$, e assim adiante. Quanto mais o tempo avança, maior será o desequilíbrio.

As soluções possíveis para lidar com este problema são as seguintes:

1. Compensar o desequilíbrio com a migração externa. Isso é possível, mas tem o inconveniente de modificar a migração externa por motivos que em realidade não têm nenhuma relação com ela.
2. Compensar o desequilíbrio com uma região, geralmente a maior. Neste caso, isto significa que os 60 migrantes que sobraram entre o ano 1 e 2 teriam que ser subtraídos do total de 1.300 da região A, de modo que haveria um saldo de apenas 1.240, em vez de 1.300, nessa região. Nos outros anos se procederia da mesma maneira.
3. Em vez de manter as TLMs constantes, se pode manter o saldo migratório de cada região constante. O inconveniente desta solução é que, neste caso, implica TLMs decrescentes para a região A e crescentes para a região C, sem justificativa. De uma forma mais geral, pode ser melhor formular as hipóteses de migração na forma de saldos líquidos consistentes, em vez de taxas inconsistentes.
4. Somar todos os migrantes que saem de regiões com saldos negativos e distribuí-los entre as regiões com saldos positivos, proporcionalmente aos tamanhos dos saldos inicialmente esperados. No exemplo acima, como há apenas uma região (A) com saldo esperado

positivo, esta solução acaba sendo igual a 2). Entretanto, quando há várias regiões com saldos iniciais positivos, esta solução tem a vantagem de diluir a correção entre elas, em vez de compensar o desequilíbrio com apenas uma delas.

21.6.3 Nascimentos sobreviventes

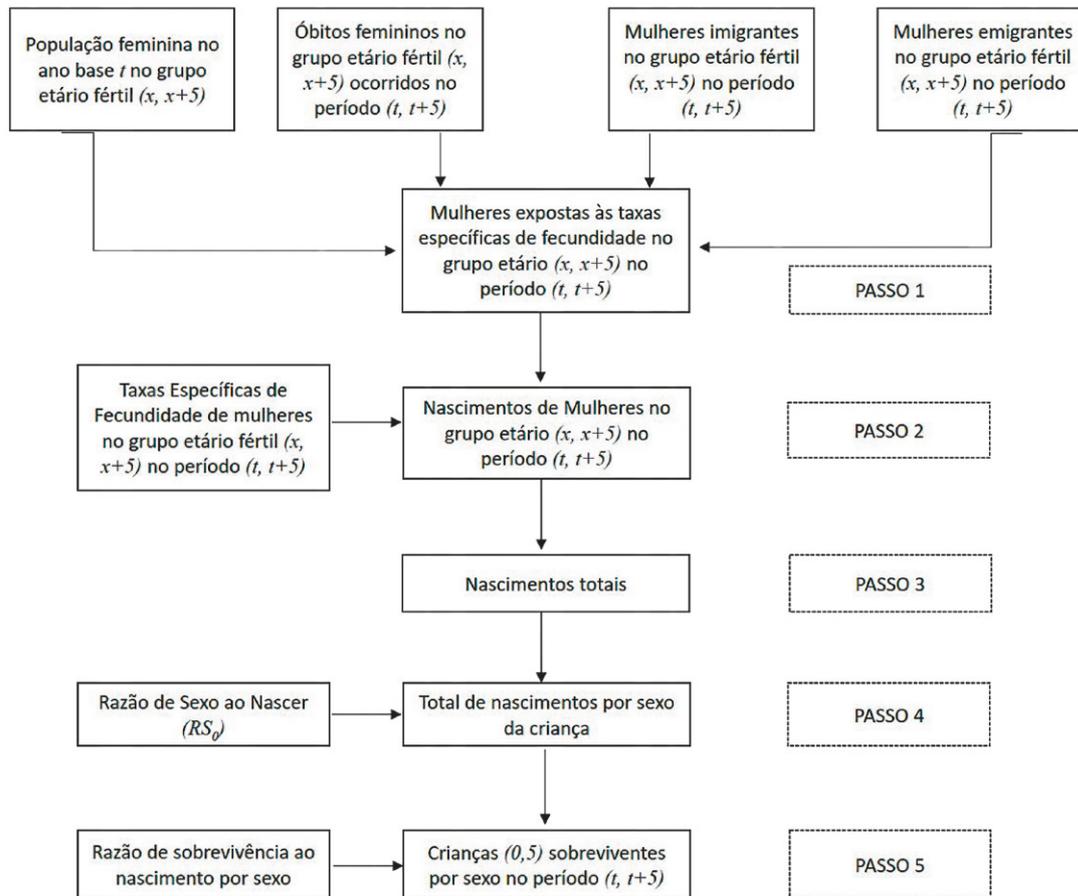
Ao usar o método das componentes, aplicando a função de mortalidade de um ano ao outro tem-se a projeção da população a cada idade sobrevivente ao grupo etário seguinte. Ou seja, sabendo-se a população entre as idades x e $x+n$ em t , ${}_n P_x(t)$, é possível estimar a população sobrevivente deste grupo etário no tempo $t+n$, ${}_n P_x(t+n)$. Assim, tem-se a população de todos os grupos etários a partir da idade $x+n$. Mas, como estimar, em tempos futuros, a população do primeiro grupo etário ${}_n P_0(t+n)$?

Essa população ${}_n P_0(t+n)$ não pertencera a essa população no tempo t , mas é fruto dos nascimentos ocorridos entre o tempo t e $t+n$. É necessário, portanto, conhecer a função de fecundidade das mulheres da população de interesse para, então, estimar o número de nascimentos dessas mulheres entre t e $t+n$ e, por meio deles, o número de nascimentos sobreviventes no tempo $t+n$. A fecundidade é a principal responsável por mudanças na estrutura etária da população, pois afeta diretamente a base dessa estrutura. Considerando a projeção para grupos etários quinquenais, nesta seção se estimará, portanto, a população ${}_5 P_0(t+5)$.

Uma vez obtida a população fechada e os migrantes (estágios 2 e 3), é que procede-se ao cálculo dos nascimentos sobreviventes no quinquênio. Ou seja, contabilizam-se os nascimentos sobreviventes de mulheres naturais da região e os nascimentos sobreviventes frutos dos efeitos diretos e indiretos das mulheres imigrantes (para mais informação sobre os efeitos indiretos da migração, ver seção 11.5.3 do Capítulo 11). A população base para o cálculo desses nascimentos é a população feminina aberta (mulheres migrantes e não migrantes sobreviventes no quinquênio) no intervalo etário reprodutivo, ou seja, população feminina em idade fértil (estágio 4).

O algoritmo para contabilizar o número de nascimentos sobreviventes é estruturado de forma a responder as seguintes questões: a) qual a contribuição das mulheres sobreviventes, em cada grupo etário do período reprodutivo, para o total de nascimentos no período ? Desses nascimentos, quantos sobrevivem para o final do quinquênio ? Quantos são do sexo masculino e feminino ? O caminho percorrido para responder estas perguntas está representado no esquema da Figura 21.3, onde se observa-se a importância e a necessidade de se fazer a projeção por sexo. Uma vez que o método contabiliza incrementos e decrementos por coortes etárias iniciadas no ano base, os nascimentos femininos sobreviventes irão compor, no médio prazo (cerca de 15 a 40 anos depois), a população feminina exposta às taxas de fecundidade projetadas, ou seja, aquela responsável pelos nascimentos da próxima geração.

Figura 21.3: Esquema de implementação da componente de fecundidade



Fonte: Elaboração própria.

Para descrever esse algoritmo de cálculo dos nascimentos sobreviventes, considere que ${}_5N_x(t,t+5)$ representa o número de nascimentos entre t e $t+5$, de mulheres no grupo etário $(x,x+5)$. Então, a TEF nesse grupo etário será:

$${}_5f_x(t, t + 5) = \frac{{}_5N_x(t, t + 5)/5}{{}_5P_x^f(t, t + 5)} \quad (21.22)$$

onde são as mulheres no grupo etário $(x,x+5)$ expostas às taxas de fecundidade no período quinquenal de projeção (Passo 1 do esquema apresentado na Figura 21.3).

Então, a contribuição das mulheres migrantes e não migrantes, no grupo etário $(x,x+5)$, para os nascimentos durante o quinquênio (segundo Passo do esquema apresentado na Figura 21.3), pode ser aproximada por $2,5 ({}_5P_x^f(t) + {}_5P_x^f(t+5)) f_x(t,t+5)$.

A multiplicação por 5 ocorre porque a estimativa é para o quinquênio e não para um ano específico, adotando-se o pressuposto de que as taxas anuais de fecundidade são homogêneas ao longo do quinquênio. É de notar que ${}_5P_x^f(t+5)$ foi projetada nos dois passos anteriores e portanto

incorpora os efeitos da mortalidade (mulheres que morreram antes de $t+5$) e da migração (mulheres que entraram ou saíram da população durante o período $(t, t+5)$).

Somando as contribuições de todas as mulheres de idade fértil, se obtém

$$N(t, t + 5) = 2,5 \sum_{x=15}^{49} ({}_5P_x^f(t) + {}_5P_x^f(t + 5)) \cdot {}_5f_x(t, t + 5) \quad (21.23)$$

onde N se refere ao total de nascimentos e a soma é executada em intervalos de 5 anos.

Desse total de nascimentos deve-se contabilizar quantos são do sexo feminino e quantos são dos sexo masculino. Assim, no que segue, resta separar esse total de nascimentos sobreviventes por sexo (o quarto passo da Figura 21.3), uma vez que os nascimentos femininos serão expostos a riscos de morte feminina durante o ciclo de vida e, ao alcançarem o período reprodutivo, estarão expostos às taxas específicas de fecundidade projetadas. A parcela dos nascimentos sobreviventes do sexo feminino é determinada utilizando-se da razão de sexo ao nascer (RS_0) (ver Capítulos 6 e 10). Assim, tem-se que o número de nascimentos sobreviventes do sexo masculino seja determinado por:

$$N^m(t, t + 5) = 2,5 \frac{RS_0}{1 + RS_0} \sum_{x=15}^{49} ({}_5P_x^f(t) + {}_5P_x^f(t + 5)) \cdot {}_5f_x(t, t + 5) \quad (21.24)$$

Já os nascimentos do sexo feminino podem ser obtidos multiplicando-se a equação (21.23) pelo complemento de $RS_0/(1+RS_0)$, que é $1/(1+RS_0)$.

Os nascimentos por sexo obtidos pela equação (21.24) (o quarto Passo da Figura 21.3) estão expostos a riscos diferenciados de mortalidade na infância de forma que é preciso subtrair da equação (21.24) somente os sobreviventes. Portanto, para obter ${}_5P_0(t+5)$ é preciso multiplicar a versão masculina e feminina de (21.24) pela razão de sobrevivência ${}_nL_x / \ell_0$ correspondente a cada sexo. Ou seja, no caso masculino:

$${}_5P_x^m(t, t + 5) = 0,5 \frac{{}_5L_x^m}{\ell_0} \frac{RS_0}{1 + RS_0} \sum_{x=15}^{49} ({}_5P_x^f(t) + {}_5P_x^f(t + 5)) \cdot {}_5f_x(t, t + 5) \quad (21.25)$$

21.6.4 Abordagem matricial do Método das Coortes Componentes

As projeções pelo métodos das componentes podem ser representadas sinteticamente pelo uso da chamada *matriz de Leslie* (Leslie, 1945). Supondo que a projeção é definida em termos de m faixas etárias quinquenais (ou de idade simples) e dois sexos, a população em cada ano da projeção pode ser escrita como um vetor de comprimento $2 \cdot m$. A transformação da população no ano t para a população no ano $t+n$ pode ser feita por meio desta matriz, que tem uma dimensão de $2 \cdot m$ por $2 \cdot m$. Em realidade, a maior parte da matriz consiste de zeros. Os únicos elementos diferentes de zero são aqueles marcados com “F” (de “Fecundidade”) ou “S” (de “Sobrevivência”) no esquema da Figura 21.4.

Ou, acrescentando uma componente de migração

$$P(t+n) = L P(t) + M(t,t+n) \quad (21.27)$$

A matriz L não pode ser invertida porque em princípio há muitas matrizes de 34 por 34 que satisfazem a equação (21.26) ou (21.27). Entretanto, ela tem uma outra propriedade importante que é a seguinte: Para valores grandes de k , a aplicação repetida de L dá

$$L^{k+1} P(t) \approx e^{r \cdot n} L^k P(t) \quad (21.28)$$

ou seja, depois de um período inicial de transição de mais ou menos k iterações, a multiplicação matricial da população por L vira equivalente à multiplicação de todas as faixas etárias da população por um fator de expansão uniforme $e^{r \cdot n}$. Na língua da álgebra matricial, $L^k P(t)$ é aproximadamente um *autovetor* de L e $e^{r \cdot n}$ é o seu *autovalor* associado. Esta propriedade é conhecida como o *princípio ergódico forte*. Ela é fundamental para a teoria de populações estáveis que será discutida no Capítulo 22.

O modelo multirregional ou de múltiplos estados generaliza a matriz da Figura 21.4 para tábuas de vida de múltiplos estados. Isso simplesmente significa que cada elemento da matriz mostrada na Figura 21.4 passa a ser uma submatriz de N por N , onde N é o número de regiões ou estados relevantes. Os elementos 0 passam a ser matrizes de N por N com apenas zeros. Os elementos S passam a ser razões de sobrevivência multidimensionais, conforme explicado na seção 19.3 do Capítulo 19. Os elementos F passam a referir-se a filhos de mulheres que no início do período tinham entre x e $x+n$ anos e moravam na região i e cujos filhos sobreviventes (homens ou mulheres, separadamente) de 0-4 anos no final do período moram na região j . O resultado é uma matriz de projeção muito grande, de $2 \cdot m \cdot N$ por $2 \cdot m \cdot N$, cuja maioria de elementos é zero. Esta matriz, quando multiplicada pelo vetor de população por sexo, idade e região ou estado no momento t , dá como resultado a população por sexo, idade e região ou estado no momento $t+n$. Se há migração externa, é preciso acrescentar um vetor $M(t,t+n)$, como em (21.26).

Com a exceção de alguma limitação ou disponibilidade de dados, nada impede que as informações acima sejam desagregadas por alguma característica de interesse e que o método seja implementado levando-se em conta essa desagregação. Por exemplo, pode ser de interesse a projeção da população por raça, escolaridade etc. Neste caso, é necessário que os dados requeridos acima estejam também desagregados pela característica de interesse. Sugere-se a leitura de Smith, Tayman e Swanson (2001) para alguns exemplos. Outra informação importante, como já mencionado, é a conveniência de se fazer a projeção por períodos quinquenais de tempo ($n=5$) ou, ao menos, no mesmo intervalo de tempo dos grupos etários.

O uso de uma metodologia de múltiplos estados pode modificar significativamente os resultados de uma projeção, como pode ser ilustrado com o exemplo das projeções demográficas estratificadas por educação. As entidades principais que preparam projeções demográficas para todos os países do mundo são a Divisão de População das Nações Unidas, o US Bureau of the Census dos EUA e o IIASA da Áustria. As projeções feitas pelas primeiras duas entidades não são

muito diferentes. Por exemplo, a Divisão de População das Nações Unidas, na sua projeção média feita em 2019, prevê um aumento da população mundial para 9.735 milhões em 2050 e um máximo de 10.880 milhões em 2103. O US Bureau of the Census projeta 9.488 habitantes para 2050 e não fornece projeções além dessa data.

Entretanto, o IIASA usa uma metodologia que projeta simultaneamente as tendências de população e educação, com as suas interações nos dois sentidos². Rodrigues, Moreira e Rodrigues (2013) aplicaram essa mesma metodologia em Portugal para projetar o nível de educação da população portuguesa até 2031. Devido ao aumento esperado da fecundidade e do efeito inibidor do aumento da educação sobre a fecundidade (e em menor medida a mortalidade), a versão mais recente projeção média do IIASA (Lutz et al., 2018) é bem mais conservadora do que as outras duas, prevendo uma população mundial máxima de 9.800 milhões na década de 70 deste século e um lento declínio para 9.500 milhões em 2100. A Tabela 21.1 mostra a diferença dos resultados para os países de língua portuguesa. Enquanto as diferenças no caso dos países com maior nível de desenvolvimento, como Portugal e Macau, não são muito significativas, elas são muito grandes nos PALOP. Um detalhe interessante é que o Brasil e Timor-Leste são os únicos países onde as projeções do IIASA *superam* as das Nações Unidas.

Tabela 21.1: Populações dos países de língua portuguesa em 2050 e 2100 (em milhões) segundo as projeções médias das Nações Unidas (que não fazem uso explícito das tendências da educação) e do IIASA (baseadas na projeção simultânea da educação e da população)

	2050		2100	
	Nações Unidas	IIASA	Nações Unidas	IIASA
Angola	77,42	58,94	188,28	76,88
Brasil	228,98	236,33	180,68	201,09
Cabo Verde	0,68	0,65	0,60	0,50
Guiné-Bissau	3,56	2,84	5,71	2,72
Guiné Equatorial	2,82	2,19	4,51	2,71
Macau	0,84	0,83	1,01	0,99
Moçambique	65,31	50,76	123,65	58,73
Portugal	9,09	8,79	6,99	7,30
São Tomé & Príncipe	0,39	0,29	0,71	0,25
Timor-Leste	2,02	2,45	2,37	2,87

Fontes: Divisão de População das Nações Unidas, Revisão de 2019; Lutz et al. (2018).

Além das projeções referidas acima, vale ainda mencionar a projeção que o Institute for Health Metrics and Evaluation (IHME), da Universidade de Washington, lançou em 2020 e que, da mesma forma que o IIASA, prevê uma estabilização mais rápida da população mundial, em 2064, num nível de 9,73 bilhões de pessoas (Vollset et al., 2020). Essa projeção se baseia no suposto de que não apenas a educação aumentará, mas que a totalidade dos objetivos dos ODS será atingida. Entretanto, este resultado vem sendo questionado por muitos profissionais da área, devido a alguns supostos não consensuais.

² Não toda a diferença entre as projeções vem por conta do efeito de educação nas projeções do IIASA. Também existem diferenças na estimação da fecundidade recente na África e na China.

21.7 MÉTODOS PARA PROJEÇÕES DE PEQUENAS ÁREAS

A projeção de populações municipais e de outras populações de pequenas áreas é de grande importância para a implementação, acompanhamento e avaliação de políticas públicas (Waldvogel, 1998). Muitos dos métodos usados para a projeção de pequenas áreas não são propriamente demográficos, como os métodos estruturais mencionados anteriormente. Entretanto, dentro dos métodos demográficos, dois grupos se destacam: o dos métodos em que primeiro se projeta a população da área maior para depois distribuí-la nas populações das áreas menores (“top down”), e o dos métodos em que faz-se o contrário, projetando primeiro as áreas menores de forma que a soma das populações das áreas menores seja equivalente à população da área maior (“bottom up”).

A ênfase nesta seção será sobre os métodos do primeiro tipo, onde se distribui uma projeção para uma área maior entre áreas menores. Os métodos principais nessa categoria são os da distribuição do crescimento (“apportionment”, em inglês) e o método das razões, exemplificado pelo método logístico da participação da população da área menor na população da área maior. A diferença entre os dois é que o primeiro distribui o *crescimento* da população enquanto o segundo distribui a própria *população* entre as áreas menores. No segundo caso, essa distribuição passa pela estimação das participações ou razões de proporcionalidade entre cada área menor e a área maior.

Ainda é preciso distinguir entre métodos que projetam apenas a população total e métodos que detalham os resultados por idade e sexo. O detalhamento pode ser introduzido de forma externa ou pode ser parte do próprio processo de projeção. Os métodos que internalizam a projeção da estrutura por idade e sexo às vezes são chamados métodos de *quasecomponentes*. O exemplo mais conhecido dessa segunda categoria é o método de Duchesne que será discutido mais adiante.

21.7.1 Métodos de distribuição do crescimento

Entre os métodos mais simples para projetar as populações de pequenas área quando o objetivo é apenas projetar a população total, sem detalhe de idade e sexo, estão as fórmulas de distribuição do crescimento, conhecidas em inglês como métodos “shift share”. Embora existam diferentes variantes desses métodos, de longe o mais conhecido é a fórmula proporcional de distribuição proposta por Pickard (1967) com o nome de “apportionment method” e introduzido no Brasil por Madeira e Simões (1972). González e Torres (2012) se referem ao método como “método de los incrementos relativos”. No Brasil, geralmente é conhecido como o método de *tendência do crescimento* ou $a_i b_i$, mas aqui será referido como o método de *distribuição* (ou *partição*, na terminologia de Santos e Barbieri, 2015) *constante do crescimento* ou simplesmente *distribuição* (*partição*) *do crescimento*, o qual é uma tradução mais fiel do termo “apportionment”³. Como será visto mais adiante, se trata de um método que, em certas situações, pode dar resultados pouco plausíveis, mas é muito fácil de aplicar e desde que a tendência de crescimento de todas as áreas projetadas seja na mesma direção e não se preveja uma reversão das tendências no futuro, os resultados tendem a ser bastante estáveis. Foi por esta razão que o IBGE o usou durante muito tempo para projetar as populações municipais no Brasil.

³ Este termo não deve ser confundido com o termo “distribuição (partição) constante” que se refere ao suposto de constância da distribuição da *população* (não do *crescimento*) entre as áreas menores.

Como foi indicado acima, parte-se do princípio de que as áreas menores fazem parte de uma área maior cuja população foi projetada previamente, usando o método dos componentes ou algum outro método apropriado. Uma das vantagens do método de distribuição do crescimento é que os resultados para as áreas menores são automaticamente consistentes com o total da área maior (Frias, 1987). Isso dispensa a necessidade de ajustes finais do tipo que, como se verá mais adiante, são necessários no caso do método logístico e outros métodos de razões. A população projetada da área maior no momento t é $P(t)$. A área maior é dividida em n áreas menores, de forma que a população da área maior seja igual à soma das populações $P_i(t)$ das áreas menores no mesmo tempo t :

$$P(t) = \sum_{i=1}^n P_i(t) \quad (21.29)$$

Então, as populações $P_i(t)$ podem ser projetadas para o tempo t_p com a seguinte fórmula:

$$P_i(t_p) = P_i(t + \Delta t) + (P(t_p) - P(t + \Delta t)) \frac{P_i(t + \Delta t) - P_i(t)}{P(t + \Delta t) - P(t)} \quad (21.30)$$

Intuitivamente, esta fórmula pode ser entendida da seguinte forma. O termo $P(t_p) - P(t + \Delta t)$ é o crescimento da área maior entre $t + \Delta t$ e o momento t_p da projeção. A fração final é a proporção do crescimento total ocorrido entre t e $t + \Delta t$ que correspondeu à área i . Supõe-se, então, que esta mesma fração se aplicará de novo no período de $t + \Delta t$ até t_p .

O mesmo resultado pode ser obtido da seguinte forma. Assumindo que, durante o período de projeção a população de cada área menor varia linearmente com a população da área maior, tem-se que (Frias, 1987):

$$P_i(t_p) = a_i P(t_p) + b_i \quad (21.31)$$

onde:

- a_i é o coeficiente de proporcionalidade do incremento da população da área menor i em relação ao incremento da população da área maior;
- b_i é o coeficiente linear de correção.

Esta mesma relação também se aplica nos momentos t e $t + \Delta t$, para cada área menor i . Além disso, para que o sistema de equações seja consistente, é preciso que

$$\sum_{i=1}^n a_i = 1 \quad \text{e} \quad \sum_{i=1}^n b_i = 0 \quad (21.32)$$

Resolvendo todas estas relações, obtém-se (21.30). Embora esta forma de implementação do método seja mais complicada do que (21.30), foi o formato usado no artigo de Madeira e Simões, a partir do qual o método ficou conhecido no Brasil como $a_i b_i$. Segundo Santos e Barbieri (2015), uma vantagem de usar a equação (21.31) é a interpretação dos seus valores e a

possibilidade de ajuste por um grupo de especialistas, conjugando a esta técnica outras subjetivas. Entretanto, não parece haver razão porque os mesmos ajustes não possam ser aplicados à forma (21.30) do método.

Apesar da sua simplicidade, o método de distribuição constante do crescimento costuma dar bons resultados em situações onde as populações de todas as áreas menores evoluem na mesma direção e onde não há mudanças bruscas no crescimento da área maior. Entretanto, uma limitação do método é que ele não é consistente quando o crescimento populacional das áreas menores apresenta direções opostas, ou seja, quando algumas áreas menores crescem enquanto outras decrescem. Também pode gerar resultados implausíveis quando o sentido de crescimento da área maior muda de positivo para negativo ou vice-versa. Finalmente, em certas situações o método pode produzir populações negativas. Uma forma de contornar a primeira restrição do método é dividir a área maior em dois subgrupos: um composto pelas áreas menores que crescem em população e outro composto pelas áreas menores que diminuem em população. Depois disso, o método deve ser aplicado separadamente para cada um desses subgrupos.

Uma outra maneira de contornar as limitações da variante de distribuição constante do crescimento é por meio de algumas fórmulas de distribuição alternativas, como a seguinte, da autoria de Van Lohuizen-Delfgaauw (1935), que inclusive é mais antiga do que a fórmula de Pickard:

$$P_i(t + 2\Delta t) = P_i(t + \Delta t) + (P_i(t + \Delta t) - P_i(t)) + \frac{P_i(t + \Delta t)}{P(t + \Delta t)} ((P(t + 2\Delta t) - P(t + \Delta t)) - (P(t + \Delta t) - P(t))) \quad (21.33)$$

Esta fórmula poderia ser caracterizada como a *fórmula de distribuição proporcional da mudança do crescimento*. O método não possui as limitações do método de distribuição constante, porém tem um outro inconveniente. O método de distribuição constante (21.30) pode ser aplicado sequencialmente para diferentes períodos ou então diretamente para o período final, sem passar pelas diferentes etapas, com o mesmo resultado. Ou seja, se for aplicado para 2010-2020, usando 2000-2010 como período de base, e depois para 2020-2030, usando 2010-2020 como período de base, ou alternativamente para 2010-2030, sem o passo intermediário, o resultado será igual. O mesmo não acontece com (21.33) cujo resultado depende da sequência de intervalos de tempo escolhida.

Abaixo se mostra um exemplo de uma situação em que as populações de algumas áreas menores crescem e outras diminuem e como as duas estratégias alternativas se aplicam a esse caso. Trata-se dos oito municípios que compõem a microrregião de Angicos, no Rio Grande do Norte (RN), Brasil. A Tabela 21.2 os separa em municípios com populações crescentes (entre 1991 e 2000) e decrescentes.

Tabela 21.2: Municípios com populações crescentes e decrescentes na microrregião de Angicos (RN) em 1991 e 2000 projetados para 2010 com os métodos de Distribuição Constante do Crescimento (Pickard) e Distribuição Proporcional da Mudança de Crescimento (Van Lohuizen-Delfgaauw), supondo uma população total de 49.711 em 2010

		1991	2000	2010	
				DCC	DPMC
Crescentes	Afonso Bezerra	10.733	10.867	11.022	11.205
	Angicos	11.535	11.626	11.731	11.935
	Caiçara do Rio do Vento	2.616	2.867	3.158	3.172
	Jardim de Angicos	2.439	2.670	2.938	2.951
	Lajes	8.687	9.399	10.224	10.287
	Pedra Preta	2.710	2.847	3.006	3.037
	Total	38.720	40.276	42.078	42.587
Decrescentes	Fernando Pedroza	2.941	2.650	2.414	2.409
	Pedro Avelino	11.447	8.006	5.219	4.715
	Total	14.388	10.656	7.633	7.124
Total Geral		53.108	50.932	49.711	49.711

NB: Os números sublinhados são dados; os demais são calculados.

Fonte: PNUD. *Atlas do Desenvolvimento Humano do Brasil*.

As primeiras duas colunas da Tabela 21.2 mostram simplesmente as populações municipais em 1991 e 2000, mais os totais dos municípios com populações crescentes e decrescentes e o total geral. O número 49.711, que é o total geral para 2010, é a população projetada da microrregião para 2010. Aqui não importa qual foi o método usado para obter esse número. Para aplicar o método de distribuição proporcional da mudança de crescimento, esta é toda a informação necessária para derivar as projeções para as populações municipais em 2010. Para aplicar o método de distribuição constante do crescimento, é preciso, além disso, separar o total geral de 49.711 nas partes que correspondem aos municípios com populações crescentes e decrescentes. Recomenda-se fazer uma projeção dos municípios com populações decrescentes e calcular o outro grupo por diferença. Aqui o número 7.633 foi obtido por meio de uma extrapolação exponencial simples. Uma vez separados os grupos, (21.30) foi aplicado separadamente para os primeiros seis municípios e para os dois municípios restantes. Os resultados são mostrados na Tabela 21.3.

Tabela 21.3: Municípios com populações crescentes e decrescentes na microrregião de Angicos RN em 1991 e 2000 projetados para 2010 com os métodos de Distribuição Constante do Crescimento (Pickard) e Distribuição Proporcional da Mudança de Crescimento (Van Lohuizen-Delfgaauw), supondo uma população total de 45.000 em 2010

		1991	2000	2010	
				DCC	DPMC
Crescentes	Afonso Bezerra	10.733	10.867	10.616	10.200
	Angicos	11.535	11.626	11.456	10.860
	Caiçara do Rio do Vento	2.616	2.867	2.398	2.906
	Jardim de Angicos	2.439	2.670	2.238	2.704
	Lajes	8.687	9.399	8.068	9.418
	Pedra Preta	2.710	2.847	2.591	2.774
	Total	38.720	40.276	37.367	38.862
Decrescentes	Fernando Pedroza	2.941	2.650	2.414	2.163
	Pedro Avelino	11.447	8.006	5.219	3.975
	Total	14.388	10.656	7.633	6.138
Total Geral		53.108	50.932	45.000	45.000

NB: Os números sublinhados são dados; os demais são calculados.

Fonte: PNUD. *Atlas do Desenvolvimento Humano do Brasil*.

Embora o procedimento de dividir as áreas menores em grupos homogêneos resolva o problema de crescimento positivo em algumas e negativo em outras, para a aplicação do método da divisão constante do crescimento, não resolve o problema de reversões na tendência de crescimento da área maior ou nos subgrupos. Na Tabela 21.1 as projeções de ambos os subgrupos mostram uma continuação da mesma tendência nos períodos de 1991-2000 e 2000-2010: o primeiro subgrupo continua a crescer, enquanto o outro subgrupo continua a diminuir. Mas se o total projetado para a microrregião é diminuído de 49.711 para 45.000, como na Tabela 21.2, o primeiro subgrupo também registra uma diminuição em 2000-2010, o que significa uma reversão da tendência de 1991-2000. Nestas condições, as projeções no primeiro subgrupo se tornam irrealistas, na medida em que os municípios que mais cresceram no período base (Caiçara do Rio do Vento e Jardim dos Angicos) acabam sendo os que perdem população mais rapidamente no período de 2000-2010. Os resultados do método de divisão proporcional da mudança de crescimento neste caso são mais plausíveis.

21.7.2 Métodos de razões

Ainda supondo que o objetivo é apenas a projeção da população de áreas menores dentro de uma maior previamente projetada, a segunda categoria de métodos de distribuição calcula a participação da população de cada área menor nos anos base e projeta essas razões. No ano da projeção usa-se o total projetado da área maior para outra vez converter as razões projetadas em números de população. A menos que as razões sejam extrapoladas ou interpoladas de forma linear, um problema que surge é que as razões projetadas normalmente não somam a 1. Por isso é preciso

fazer uma normalização que reestabelece a consistência dos resultados, de modo que o total das áreas menores equivale à população projetada da área maior.

O método linear de razões (também chamado “shift-share method”, em inglês) tem a seguinte forma:

$$P_i(t_p) = P(t_p) \left(\frac{P_i(t + \Delta t)}{P(t + \Delta t)} + \frac{t_p - t - \Delta t}{\Delta t} \left(\frac{P_i(t + \Delta t)}{P(t + \Delta t)} - \frac{P_i(t)}{P(t)} \right) \right) \quad (21.34)$$

A especificação linear, de uma forma geral, não é recomendada porque ela tende a exagerar as tendências divergentes entre áreas menores. Particularmente, (21.34) pode levar a resultados negativos ou a áreas menores com mais população do que a totalidade da área maior.

Um caso particular é a extrapolação ou interpolação exponencial ou geométrica que pode ser aplicada de duas formas equivalentes. Por um lado, é possível projetar as *razões* exponencialmente para o ano de projeção, dividir o resultado pela soma das razões para normalizá-la e depois multiplicar pela população da área maior. Entretanto, um resultado idêntico é obtido ao aplicar uma projeção exponencial ou geométrica convencional, como em (7.2.a), das *populações* das áreas menores, normalizando o resultado diretamente para obter a consistência com a população da área maior, sem passar pelo cálculo de razões:

$$P_i(t_p) = P(t_p) P_i(t) (1 + r_i)^{t_p - t} / \sum_{i=1}^n P_i(t) (1 + r_i)^{t_p - t} \quad (21.35)$$

A extrapolação exponencial é a única que possui essa propriedade; em todos os outros casos, o método das razões dá resultados diferentes da extrapolação ou interpolação direta das populações.

Uma das variantes mais usadas dos métodos de razões é o método logístico que evita uma característica indesejável do método exponencial. No método exponencial, no longo prazo a participação da área menor com o crescimento mais rápido sempre tende a 100%, de modo que as outras áreas desaparecem. Usando a especificação logística, esse problema pode ser evitado, atribuindo uma proporção inferior I_i e superior S_i ao tamanho da área menor i em relativo à área maior. Seguindo os mesmos passos descritos na seção 21.2, a proporção $P_i(t)/P(t)$ primeiro é normalizada e depois linearizada usando a transformação logito. Usando $p_i(t)$ para referenciar a razão normalizada, o resultado é o seguinte:

$$\text{logito}(p_i(t)) = \frac{1}{2} \ln \left(\frac{S - I}{P_i(t)/P(t) - I} - 1 \right) \quad (21.36.a)$$

$$\text{logito}(p_i(t + \Delta t)) = \frac{1}{2} \ln \left(\frac{S - I}{P_i(t + \Delta t)/P(t + \Delta t) - I} - 1 \right) \quad (21.36.b)$$

$$\text{logito}(p_i(t_p)) = \frac{(t + \Delta t - t_p) \text{logito}(p_i(t)) + (t_p - t) \text{logito}(p_i(t + \Delta t))}{\Delta t} \quad (21.36.c)$$

$$P_i(t_p) = \left(I + \frac{S - I}{1 + \exp(2 \text{logito}(p_i(t_p)))} \right) P(t_p) \quad (21.36.d)$$

Entretanto, não há nenhuma garantia de que a soma das populações $P_i(t_p)$ assim derivadas seja igual à população $P(t_p)$ da área maior⁴. Por isso, é preciso fazer um ajuste proporcional final, para forçar a consistência dos resultados⁵:

$$P_i^{corr.}(t_p) = P_i(t_p) \frac{P(t_p)}{\sum_{i=1}^n P_i(t_p)} \quad (21.37)$$

Na Tabela 21.4 o procedimento é aplicado ao caso da microrregião de Angicos RN que também foi analisada nas Tabelas 21.1 e 21.2.

Tabela 21.4: Municípios com populações da microrregião de Angicos (RN) em 1991 e 2000 projetados para 2010 com o método de logístico de razões

Município	1991	2000	I	S	(1)	(2)	(3)	(4)	(5)	(6)	(7)	2010
Afonso Bezerra	10733	10867	0,02	0,84	0,222	0,236	0,627	0,588	0,545	0,252	0,226	10973
Angicos	11535	11626	0,02	0,84	0,240	0,254	0,575	0,539	0,498	0,270	0,241	11680
Caiçara R. do Vento	2616	2867	0,02	0,84	0,036	0,044	1,648	1,536	1,412	0,056	0,066	3198
Fernando Pedroza	2941	2650	0,02	0,84	0,043	0,039	1,550	1,601	1,659	0,035	0,049	2358
Jardim de Angicos	2439	2670	0,02	0,84	0,032	0,040	1,711	1,595	1,466	0,051	0,061	2979
Lajes	8687	9399	0,02	0,84	0,175	0,201	0,775	0,691	0,598	0,232	0,210	10198
Pedra Preta	2710	2847	0,02	0,84	0,038	0,044	1,618	1,542	1,457	0,051	0,062	3012
Pedro Avelino	11447	8006	0,02	0,84	0,238	0,167	0,581	0,802	1,049	0,109	0,110	5313
Total	53108	50932									1,026	49711

Fonte: PNUD. *Atlas do Desenvolvimento Humano do Brasil*.

As primeiras colunas da Tabela 21.4 mostram as populações de 1991 e 2000 e os limites inferior (I) e superior (S), que aqui foram escolhidos iguais a 0,02 e 0,84, respectivamente, para todos os municípios. Mas essa escolha é arbitrária: nada impede que se escolham limites diferentes para cada município. As colunas (1) e (2) contêm as razões normalizadas $p_i(1991)$ e $p_i(2000)$. Nas colunas (3) e (4) essas razões são transformadas pelo logito como em (21.35.a) e (21.35.b). Na coluna (5) se faz a extrapolação para 2010, como descrita em (21.35.c). Na coluna (6) se aplica a transformação inversa do logito e em (7) se desfaz a normalização, para voltar a razões convencionais. Como se percebe, a soma das razões da coluna (7) é 1,026, em vez de 1. Por isso, a última coluna divide todas as razões por 1,026 e multiplica por 49.711, a população da microrregião em 2010.

Embora não seja conclusivo, um método frequentemente usado para avaliar a qualidade de métodos de projeção para pequenas áreas é a comparação das populações efetivas de certas áreas para um momento recente com as projeções obtidas com base em dados anteriores (por exemplo, Waldvogel, 1998; Brito, Cavenaghi e Jannuzzi, 2010; Santos e Barbieri, 2011). A Tabela 21.5 compara as populações efetivas dos municípios da microrregião de Angicos em 2010 com as diferentes projeções preparadas com base na informação de 1991 e 2000. Os resultados evidenciam que nos

⁴ Isso não acontece só no caso da função logística. No caso da função exponencial, Keyfitz (1977: 14-18) mostra que a soma das partes projetadas sempre é maior do que a soma projetada diretamente.

⁵ Existem outras maneiras para forçar a consistência. Por exemplo, é possível multiplicar todos os limites superiores S por um fator de ajuste c , onde c é escolhido de tal forma que a soma das razões projetadas (coluna (7) na Tabela 21.4) seja igual a 1.

municípios Afonso Bezerra, Angicos, Caiçara do Rio do Vento e Pedro Avelino, o método logístico apresentou os valores mais próximos dos observados, ao passo que para os municípios Fernando Pedroza, Jardim de Angicos e Pedra Preta, o método de distribuição constante do crescimento foi mais adequado. No município de Lajes foi o método de distribuição proporcional da mudança de crescimento (DPMC) que produziu o resultado mais próximo do observado.

Tabela 21.5: População observada de cada município da microrregião de Angicos (RN) em 2010, população projetada pelos métodos de Distribuição Constante do Crescimento (DCC), Distribuição Proporcional da Mudança de Crescimento (DPMC) e pelo método logístico de razões

Município	Observada 2010	DCC	DPMC	Logística
Afonso Bezerra	10.844	11.022	11.205	10.973
Angicos	11.549	11.731	11.935	11.680
Caiçara do Rio do Vento	3.308	3.158	3.172	3.198
Fernando Pedroza	2.854	2.414	2.409	2.358
Jardim de Angicos	2.607	2.938	2.951	2.979
Lajes	10.381	10.224	10.287	10.198
Pedra Preta	2.590	3.006	3.037	3.012
Pedro Avelino	7.171	5.219	4.715	5.313
Total	51.304	49.711	49.711	49.711

Fonte: Elaboração própria a partir da Tabela 21.2.

Uma maneira mais formal para analisar a qualidade do ajuste é calcular a soma dos desvios quadrados entre as populações projetadas e observadas. Isso pode ser feito tanto para os números absolutos como para as razões $p_i(t_p)$. Em ambos os casos, a projeção logística dá o melhor resultado, seguida pela distribuição constante do crescimento, com a distribuição proporcional da mudança de crescimento em terceiro lugar. Cabe ressaltar que este é um resultado que pode não se repetir para outras populações, uma vez que, dependendo do comportamento demográfico das pequenas áreas, não será sempre o mesmo método que dará os melhores resultados. Por exemplo, o método da distribuição proporcional da mudança de crescimento (DPMC) funciona melhor em situações onde os padrões de crescimento são muito heterogêneos no tempo e entre as unidades menores.

Jannuzzi (2007), seguindo uma proposta mais antiga de Szwarcwald e Castillo (1989), sugere um modelo mais complexo baseado em razões que chama ProjPeq e que distingue entre uma componente de crescimento vegetativo a_i e uma componente migratória b_i . A racionalidade subjacente ao modelo é o suposto de que as áreas menores se comportam como se fossem espécies competindo por um mesmo espaço ecológico. Usando outra vez $p_i(t)$ para referenciar a razão normalizada da área i , as fórmulas básicas do modelo são as seguintes:

$$p_i(t_p) = 1 + F_i(t_p)P_i(t) / \sum_{i=1}^n F_i(t_p)P_i(t) \tag{21.38.a}$$

onde

$$F_i(t_p) = \exp(a_i(t_p) + b_i(t_p) \ln\left(\frac{P(t_p)}{P(t)}\right)) (P(t_p) - P(t)) - 1 \tag{21.38.b}$$

$$a_i(t) = \ln(1 + TBN_i - TBM_i) - \text{taxas médias para o período} \quad (21.38.c)$$

$$b_i(t) = \text{grau de atratividade migratória média do período} \quad (21.38.d)$$

A estimação de a_i se baseia nas estatísticas vitais do período base e sua projeção para o futuro. A atratividade migratória b_i para o período base pode ser estimada como

$$b_i(t) = \frac{a_i(t+\Delta t) + \ln\left(\frac{P_i(t+\Delta t)}{P_i(t)}\right)}{\ln\left(\frac{P(t+\Delta t)}{P(t)}\right)(P(t+\Delta t) - P(t))} \quad (21.39)$$

A aplicação do modelo requer, como um dos insumos, a trajetória dos níveis de natalidade e mortalidade em cada área menor no horizonte de projeção. Portanto, a disponibilidade de séries históricas de razoável extensão desses indicadores é uma exigência para o seu uso. Para cada área menor, se requerem hipóteses acerca da evolução, por um lado, das taxas de natalidade e mortalidade e, por outro, da atratividade migratória ou residencial. Isso restringe a sua aplicação em comparação com os modelos anteriores, que são mais simples e exigem menos dados.

21.7.3 Métodos matriciais

O método matricial mais detalhado e mais completo para projetar um conjunto de áreas é o método multirregional ou de múltiplos estados. Entretanto, a complexidade da informação necessária para aplicar este método faz com que frequentemente seja aconselhável optar por métodos simplificados, que possuam algumas das características do método multirregional, mas que também possam ser vistos como extensões de métodos de extrapolação matemática mais simples. O ponto de partida é uma simples projeção geométrica como em (21.35):

$$P_i(t_p) = c(t_p)P_i(t)(1 + r_i)^{t_p-t} \quad (21.40)$$

onde $c(t_p)$ é um fator de normalização para garantir a consistência com a projeção $P(t_p)$ do total:

$$c(t_p) = P(t_p) / \sum_{i=1}^n P_i(t)(1 + r_i)^{t_p-t} \quad (21.41)$$

Agora, se pode substituir a população $P_i(t)$ e o fator escalar de multiplicação $(1+r_i)^{t_p-t}$ por um vetor $\mathbf{P}(t)$, que reúne as populações $P_i(t)$ de todas as áreas menores, e uma matriz multiplicativa \mathbf{R} que reúne todos os fatores de crescimento r_{ij} . Desta maneira, (21.40) se transforma em

$$\mathbf{P}(t_p) = c(t_p)\mathbf{R}^{t_p-t}\mathbf{P}(t) \quad (21.42)$$

O fator $c(t_p)$ não está em negrito porque continua sendo escalar. A sua função é garantir que o total dos elementos em $P(t_p)$ seja igual à população projetada da área maior. A matriz de crescimento R tem formato de n por n . A diagonal representa o crescimento autônomo de cada área menor, ou seja, o crescimento decorrente das suas taxas de natalidade, mortalidade e a migração entre a área menor i e o mundo externo, fora da região maior. Os fatores fora da diagonal representam o crescimento, durante o período de base Δt , da área menor i devido à área menor j , ou seja, a migração de j para i , como proporção da população inicial de j . Não é impossível definir as potências de matrizes para números não inteiros, mas para que (21.42) possa ser aplicado com facilidade, é desejável que $t_p - t$ seja um múltiplo do período de base Δt , ou seja, que seja medido como 1, 2, 3, 4 etc. períodos anuais ou quinquenais.

A vantagem da representação (21.42) é que ela já contempla certos aspectos da dinâmica regional que não são captados por (21.35). Por exemplo, se uma determinada área menor i cresce muito devido à atração que ela exerce sobre outras áreas menores da mesma região maior, (21.42) já prevê que o mesmo processo não pode continuar de forma ilimitada na medida em que a área i cresce e o tamanho das outras áreas diminui absoluta ou relativamente. Estas vantagens são ilustradas em algum detalhe em Rogers (1975), entre outros. Muniz (2018) usou esta metodologia para projetar populações municipais no Brasil, em conjunto com um procedimento Bayesiano para a estimação de R .

21.7.4 Acrescentar a Estrutura por Idade e Sexo

Os métodos de pequenas áreas apresentados até agora apenas fornecem projeções das populações totais. Mas para muitas aplicações é preciso ter pelo menos uma noção básica da estrutura por idade e sexo das populações projetadas e uma projeção convencional por componentes pode não ser factível. Existem duas maneiras para providenciar essa estrutura. Uma é o *ajuste biproportional iterativo* (“iterative proportional fitting”, em inglês) que o CELADE chama de “método de tabla cuadrada”; a outra é o uso de *quasecomponentes*, ou seja, modelagens da mudança de população que se aproximam de componentes convencionais, mas que podem ser calculadas com menos informação. A forma de aplicação do ajuste biproportional é a seguinte:

1. Monta-se um quadro com a estrutura por idade e sexo nas linhas e as áreas menores nas colunas. A última linha contém as projeções das populações totais das áreas menores e a última coluna contém a estrutura por idade e sexo da área maior no ano de projeção.
2. No interior da tabela se coloca alguma aproximação razoável de uma estrutura por idade e sexo das áreas menores, por exemplo as populações das áreas menores por idade e sexo no segundo ano base.
3. Com essas especificações, a tabela geralmente será inconsistente: as linhas não somarão ao total da última coluna, nem as colunas somarão ao total da última linha. Para forçar a consistência, primeiro se faz um ajuste proporcional por linha, de modo que as linhas somem ao total mostrado na última coluna.

4. Entretanto, os totais das colunas ainda estarão inconsistentes com a última linha. Portanto, aplica-se um ajuste por coluna, para forçar a consistência vertical.
5. Agora, a tabela outra vez deixou de ser consistente no sentido horizontal, de modo que é preciso novamente ajustar as linhas.
6. Alternando várias vezes entre ajustes das linhas e das colunas, a tabela converge para uma configuração estável. Antigamente, os cálculos iterativos para encontrar a solução estável eram onerosos, mas hoje em dia é fácil montar uma planilha em EXCEL que resolve o problema. Normalmente 4 ou 5 iterações são suficientes.
7. Eventualmente pode ser usado o módulo CTBL32 do pacote PASEX que executa o procedimento de forma automática.

Quadro 21.1: Esquema da aplicação do método de ajuste biproporcional

	Área Menor 1	Área Menor 2	Área Menor n	Total
Homens	Inicialmente populações das Áreas 1, ... , n no segundo ano base desagregadas por idade e sexo Depois ajustado iterativamente para ser consistente com os totais marginais				Área Maior Homens 0-4
0-4					Área Maior Homens 5-9
5-9					Área Maior Homens 10-14
10-14				
....				
....					
80-84					
85+					
Mulheres					
0-4					Área Maior Mulheres 0-4
5-9	Área Maior Mulheres 5-9				
10-14	Área Maior Mulheres 10-14				
....				
....				
80-84					
85+					
Total	Projetado 1	Projetado 2	Projetado n	

Fonte: Elaboração própria.

Embora o procedimento do ajuste biproporcional resulte em estruturas por idade e sexo que parecem razoáveis, se trata de um método bastante arbitrário que não possui uma justificativa demográfica inerente. Por exemplo, se a população da área menor i no segundo ano base contém alguma particularidade notável, como um número excepcional de pessoas na idade de 15-19 anos, a população projetada da área i manterá essa característica embora a lógica demográfica dite que essa anomalia deveria passar adiante, para a faixa etária dos 20-24, depois 25-29 e assim por diante. Por isso, especialmente em situações onde existem irregularidades nas estruturas etárias a serem projetadas, pode ser preferível usar um método que reproduz a estrutura etária de uma forma mais orgânica. O exemplo principal desse tipo de método de quasecomponentes é o método de relação de coortes proposto por Duchesne (1989), que será discutido a seguir.

21.7.5 Método de Relação de Coortes

O método proposto por Duchesne (1989) – conhecido como método de relação de coortes – pode ser considerado um método geométrico de razões, com a diferença de que, em vez de ser aplicado à população total, o algoritmo é aplicado separadamente a cada coorte. Os dados básicos – ou iniciais – necessários são a composição da população por sexo e grupos quinquenais de idade, para as áreas menores, e projeções de população por sexo e grupos quinquenais de idade referentes à divisão geopolítica ou administrativa maior que compreenda essas áreas menores. O algoritmo varia de acordo com as coortes às quais é aplicado:

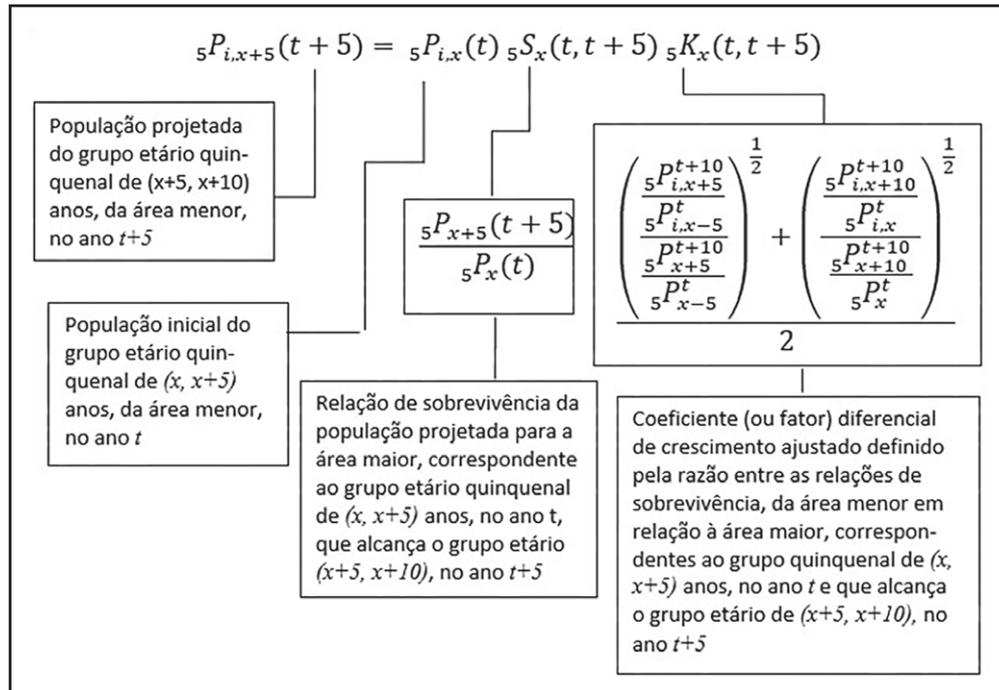
1. Cálculo da população por sexo em cada grupo quinquenal intermediário, ou seja, desconsiderando o primeiro (0-4 anos) e o último (aberto) grupo etário;
2. Cálculo dos nascimentos sobreviventes por sexo a cada período quinquenal de projeção (população no grupo etário de 0-4 anos); e
3. Cálculo da população por sexo no último grupo etário (grupo etário aberto).

Cada algoritmo é apresentado em detalhe a seguir. Da mesma forma como anteriormente, $P(t)$ se refere à população da área maior no momento t e $P_i(t)$ à população da área menor i . Além disso, serão usados os seguintes símbolos:

$N_{i,t,t+n}$	Nascimentos ocorridos na área menor i no período $(t,t+5)$;
${}_nS_x(t,t+n)$	Razão intercensitária de sobrevivência (RIS) das pessoas que tinham entre x e $x+n$ anos no momento t durante n anos, na área maior;
${}_nK_{i,x}(t,t+n)$	Índice diferencial de crescimento ao nascimento ou de crescimento das coortes na área menor i durante o período $(t,t+n)$.

Uma característica geral do método é que S se refere a *razões intercensitárias de sobrevivência* ou *razões de crescimento das coortes* (as RIS da seção 16.4) que não consideram só a mortalidade, mas também a migração. Portanto, se num determinado grupo etário a probabilidade de sobreviver por 5 anos é 0,98 e há uma migração líquida positiva de 0,5%, a razão S será igual a 0,985. A mesma mistura de mortalidade com migração também afeta os fatores K . O método de relação de coortes não identifica a migração como tal, apesar de ser uma componente importante da dinâmica de população de pequenas áreas.

Quadro 21.2: Algoritmo 1: População projetada nos grupos etários x a $x+5$ anos ($x = 5, 10, 15, \dots, 85$) para o ano $t+5$



Fonte: Elaborado com base em Duchesne (1989).

O Algoritmo 1 determina a população projetada por sexo e grupos etários quinquenais em cada área menor, pertencente a uma determinada área maior, para cada período de 5 anos. Os dados necessários para operacionalizar este algoritmo são: a) a população por sexo e grupos quinquenais de idade de cada município nos dois últimos censos demográficos (de preferência com um período intercensitário de 5 ou 10 anos); b) a população das áreas maiores projetadas por sexo, grupos etários quinquenais e anos quinquenais, tendo como ano base o ano t . O método é mais fácil de aplicar quando a distância entre os dois censos do período base for de 5 anos, como no Canadá, onde foi originalmente desenvolvido. Nesse caso, a estimação dos fatores K é direta.

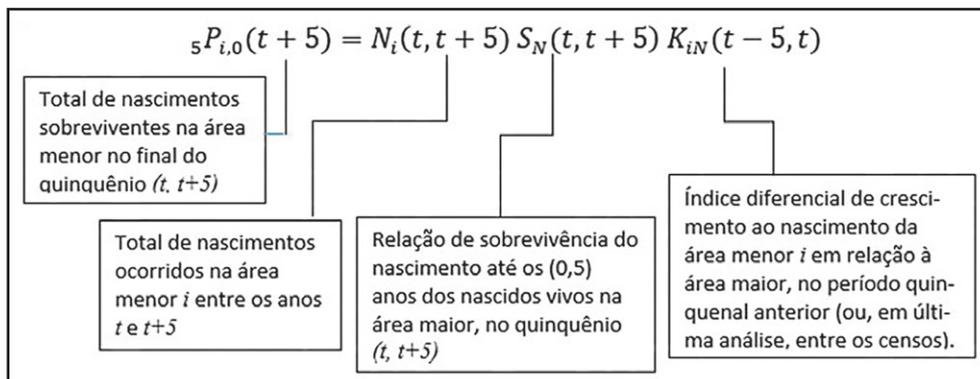
$${}_5K_x(t, t + 5) = \frac{{}_5P_{i,x+5}(t+5)}{\frac{{}_5P_{i,x}(t)}{{}_5P_{x+5}(t+5)}} \quad (21.43)$$

No caso mais comum de censos decenais é preciso desagregar os fatores K decenais em componentes quinquenais, usando a seguinte fórmula:

$${}_5K_x(t, t + 5) = \frac{1}{2} \left(\frac{\left(\frac{{}_5P_{i,x+5}(t+10)}{{}_5P_{i,x-5}(t)} \right)^{\frac{1}{2}}}{\left(\frac{{}_5P_{x+5}(t+10)}{{}_5P_{x-5}(t)} \right)^{\frac{1}{2}}} + \frac{\left(\frac{{}_5P_{i,x+10}(t+10)}{{}_5P_{i,x}(t)} \right)^{\frac{1}{2}}}{\left(\frac{{}_5P_{x+10}(t+10)}{{}_5P_x(t)} \right)^{\frac{1}{2}}} \right) \quad (21.44)$$

Em casos onde a diferença entre os censos não é de 5 ou 10 anos, o cálculo de K é mais complicado e exige intervalos de idade simples ou então a interpolação de intervalos quinquenais.

Quadro 21.3: Algoritmo 2: População projetada no grupo etário de 0-4 anos para o ano $t+5$



Fonte: Elaborado com base em Duchesne (1989)

Neste algoritmo as quantidades N e K não são diretamente observadas nos dados e S vem da projeção da área maior. A disponibilidade e baixa qualidade dos dados de população e registros vitais em áreas menores, especialmente naquelas com baixo volume de população, impossibilitam o cálculo direto desses indicadores. A solução apresentada 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 desse índice diferencial requer, inicialmente, estimativas da Razão Criança-Mulher, introduzida em (6.9) e (10.11), tanto para as áreas menores quanto para a área maior à qual pertencem. Assim, o procedimento detalhado a seguir busca suprir tais limitações nos dados e possibilita estimar indiretamente as quantidades de nascimentos sobreviventes das áreas menores de um país durante um período de tempo quinquenal.

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

$$RCM_i = \frac{{}_5P_{i,0}}{30P_{i,15}^f} \quad (21.45)$$

onde o sobrescrito “ f ” no denominador faz referência à população de mulheres no intervalo etário reprodutivo, assumindo aqui que este intervalo se inicia na idade 15 e termina aos 45 anos, ou seja, as mulheres expostas ao risco de gerar um nascimento. Da mesma forma se calcula a RCM 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 por:

$$IDF_i = \frac{RCM_i}{RCM} \quad (21.46)$$

Assume-se que o IDF_i é constante durante todo o período de projeção.

Em seguida calcula-se a Taxa de Fecundidade Total para cada área menor i equivalente ao quinquênio de projeção ($t, t+5$):

$$TFT_i(t, t + 5) = IDF_i \cdot TFT(t, t + 5) \quad (21.47)$$

Observa-se que a TFT na área menor i é o produto do Índice Diferencial de Fecundidade de cada área menor e da Taxa de Fecundidade Total da área maior.

Supõe-se que o padrão de fecundidade por idade seja o mesmo para a área maior e as áreas menores, ou seja

$${}_5TEF_{i,x}(t, t + 5) = \frac{{}_5TEF_x(t, t + 5)}{TFT(t, t + 5)} TFT_i(t, t + 5) \quad (21.48)$$

O número N_i de nascimentos na área i agora pode ser calculada da forma habitual, multiplicando cada $TEF_{i,x}$ por 5 vezes a média do número de mulheres na faixa correspondente em t e $t+5$ (este último precisa ser calculado primeiro, usando o segundo algoritmo abaixo).

Os nascimentos assim determinados estão expostos a um risco de morte desde o nascimento até antes de completarem o 5º aniversário. Ademais, a equação (21.48) não distingue entre nascimentos femininos e masculinos. O cálculo dos nascimentos sobreviventes masculinos e femininos, respectivamente, na área maior i dar-se-á por:

$${}_5P_{i,0}^m(t + 5) = N_i(t, t + 5) \frac{RS_0}{1+RS_0} \frac{{}_5L_0^m(t,t+5)}{5\ell_0^m} K_{i,N}^m(t, t + 5) \quad (21.49.a)$$

$${}_5P_{i,0}^f(t + 5) = N_i(t, t + 5) \frac{1}{1+RS_N} \frac{{}_5L_0^f(t,t+5)}{5\ell_0^f} K_{i,N}^f(t, t + 5) \quad (21.49.b)$$

onde

$$K_{i,N}(t, t + 5) = {}_5K_{i,0}(t, t + 5)^{0.5} \quad (21.50)$$

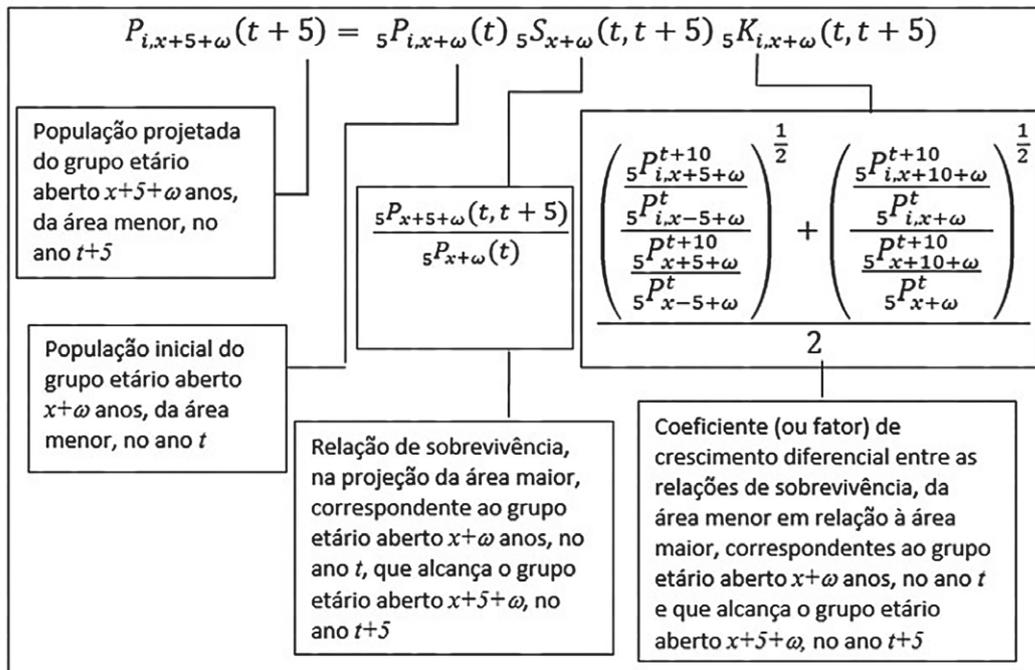
Em (21.49.a) e (21.49.b) os fatores $RS_0/(1+RS_0)$ e $1/(1+RS_0)$ determinam as parcelas de nascimentos masculinos e femininos, respectivamente. Já o respectivo fator ${}_5L_0/5\ell_0$ para cada sexo, determina o risco de sobrevivência ao nascimento na área maior e deve ser obtido com base nas tábuas de vida projetadas para a área maior. Em (21.50), ajusta-se o ${}_5K_0$ equivalente ao período de um quinquênio para um K equivalente a um ano, já que o índice diferencial de crescimento ao nascimento refere-se somente ao primeiro ano. Vale mencionar que Gonzaga, Freire e Gomes (2019) desenvolveram uma versão probabilística de (21.50) usando a metodologia bayesiana empírica.

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 (K_N tal como definido em 21.49). Primeiro, se o intervalo intercensitário é de 10 anos, não existe a informação necessária para estimar o fator ${}_5K_0$. Neste caso, recomenda-se a utilização do fator ${}_5K_5$, ou seja, do fator diferencial de crescimento ajustado. 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 atribuir ao k_b , os fatores ${}_5K_{20}$ ou ${}_5K_{25}$.

O Algoritmo 3) determina a população projetada em cada área menor, pertencente a uma determinada área maior, por sexo e grupo etário aberto (último grupo de idade) num período de 5 anos. Os dados necessários para operacionalizar este algoritmo são: a) população por sexo e grupos quinquenais de idade de cada área menor nos dois últimos censos demográficos (sendo necessário um período intercensitário de 5 ou 10 anos); b) população da área maior projetada por sexo, grupos etários quinquenais e anos quinquenais, tendo como ano base o ano t .

Quadro 21.4: Algoritmo 3: População projetada no grupo etário aberto para o ano $t+5$



Fonte: Elaboração própria com base em Duchesne (1989).

21.8 EXEMPLO PRÁTICO

O exemplo que segue aplica o método das coortes componentes para projeção da população do Estado do Rio Grande do Norte (Brasil), por sexo e grupos etários quinquenais, de 2010 para 2015. Em seguida, se projeta a população das microrregiões do Estado mediante aplicação dos métodos de Distribuição Constante do Crescimento (Pickard) e de Relação de Coortes (Duchesne, 1989). Como já argumentado, o método de Distribuição Constante do Crescimento tem sido utilizado pelo Instituto Brasileiro de Geografia e Estatística para projeções municipais no Brasil. Entretanto, o método tem suas limitações e somente fornece projeções da população total e/ou por sexo. Nesse sentido, pode-se considerar o Método de Relação das Coortes como completar ou alternativo, na medida em que fornece projeções por grupos de idade e sexo. Neste exemplo, todos os cálculos são conduzidos em “R”, detalhando-se cada etapa do procedimento.

Os dados de entrada necessários para implementação do Método das Coortes Componentes para projeção da população de 2010 para 2015 são:

1. População por sexo e idade em 2010;
2. Tábuas de mortalidade por sexo e grupos quinquenais de idade que expressem as razões de sobrevivência (mediante o uso da função ${}_nL_x$, ver definição no Capítulo 9) para o respectivo período quinquenal de projeção;
3. TEFs por grupos quinquenais de idade, que expressem o potencial reprodutivo da população no respectivo período quinquenal; e
4. Taxas líquidas de migração para o Estado do Rio Grande do Norte para o respectivo período quinquenal.

Os resultados deste exemplo prático são mostrados na Tabela 21.6.

Primeiramente, são carregados os dados de entrada mediante uso da função “read.csv”. Neste caso, todos os arquivos de dados devem estar no formato “.csv”:

```
pop = read.csv('pop2010.RN.csv', sep=';', stringsAsFactors = FALSE)
nLx = read.csv('nLx.csv', sep=';', stringsAsFactors = FALSE)
TEF = read.csv('TEF.csv', sep=';', stringsAsFactors = FALSE)
TLM = read.csv('TLM.csv', sep=';', stringsAsFactors = FALSE)
```

As estruturas dos arquivos “pop” e “nLx” devem ser exatamente como segue:

pop2010.RN.csv:			nLx.csv:			TLM.csv:		
sexo	idade	pop2010	sexo	idade	nLx	sexo	idade	TLM
fem	0	126158	fem	0	492330	fem	0	0.00363
fem	5	133256	fem	5	491516	fem	5	0.00455
fem	10	145957	fem	10	490825	fem	10	0.01500
fem	15	149374	fem	15	490061	fem	15	0.01039
fem	20	156576	fem	20	488943	fem	20	0.00034
fem	25	147147	fem	25	487345	fem	25	-0.03753
fem	30	130679	fem	30	485348	fem	30	-0.00437
fem	35	116755	fem	35	482723	fem	35	0.03438
fem	40	112973	fem	40	478966	fem	40	0.04085
fem	45	101144	fem	45	473249	fem	45	0.02134
fem	50	78679	fem	50	464310	fem	50	0.03590
fem	55	63016	fem	55	451427	fem	55	0.04756
fem	60	57135	fem	60	434068	fem	60	0.06690
fem	65	42481	fem	65	409861	fem	65	0.05402
fem	70	35305	fem	70	375431	fem	70	0.06173
fem	75	23538	fem	75	325068	fem	75	0.13123
fem	80	18305	fem	80	253868	fem	80	0.30433
fem	85	10685	fem	85	168806	fem	85	0.56211
fem	90	6639	fem	90	130791	fem	90	1.35873
masc	0	131289	masc	0	489204	masc	0	0.00358
masc	5	137770	masc	5	487942	masc	5	0.00449
masc	10	151688	masc	10	486870	masc	10	0.01361
masc	15	151873	masc	15	484228	masc	15	0.01031
masc	20	157120	masc	20	479179	masc	20	-0.01005
masc	25	144821	masc	25	472920	masc	25	-0.05353
masc	30	125795	masc	30	466420	masc	30	-0.02015
masc	35	109931	masc	35	458685	masc	35	0.03550
masc	40	105020	masc	40	448969	masc	40	0.04573
masc	45	92420	masc	45	437068	masc	45	0.02782
masc	50	69628	masc	50	421173	masc	50	0.04084
masc	55	53407	masc	55	400093	masc	55	0.05049
masc	60	47903	masc	60	372347	masc	60	0.07002
masc	65	34641	masc	65	335425	masc	65	0.08590
masc	70	27776	masc	70	289246	masc	70	0.10704
masc	75	17425	masc	75	231485	masc	75	0.15727
masc	80	13370	masc	80	162466	masc	80	0.32553
masc	85	7924	masc	85	93321	masc	85	0.62132
masc	90	4336	masc	90	55022	masc	90	1.28027

Já o arquivo “TEF.csv” deve conter as taxas específicas de fecundidade projetadas para o ponto médio de cada período quinquenal e sua estrutura deve ser como segue:

Ano	Idade	TEF
2012.5	15	0.0660
2012.5	20	0.0888
2012.5	25	0.0817
2012.5	30	0.0635
2012.5	35	0.0329
2012.5	40	0.0089
2012.5	45	0.0008

As variáveis em cada arquivo de dados de entrada devem ser configuradas como segue:

```
pop$sexo = as.character(pop$sexo)
pop$idade = as.integer(pop$idade)
pop$pop2010 = as.numeric(pop$pop2010)
```

```
nLx$sexo = as.character(nLx$sexo)
nLx$idade = as.integer(nLx$idade)
nLx$nLx = as.numeric(nLx$nLx)
```

```
TLM$sexo = as.character(TLM$sexo)
TLM$idade = as.integer(TLM$idade)
TLM$TLM = as.numeric(TLM$TLM)
```

```
TEF$ano = as.numeric(TEF$ano)
TEF$idade = as.integer(TEF$idade)
TEF$TEF = as.numeric(TEF$TEF)
```

Para projeção da população fechada sobrevivente (conforme equação 21.19) calcula-se, inicialmente, as razões de sobrevivência por sexo, grupos etários quinquenais (${}_nS_x$) e para o grupo etário aberto nas tábuas de vida projetadas. Em seguida, multiplica-se as populações por sexo, grupos etários quinquenais e grupo etário aberto, no ano base (2010), pelas respectivas razões de sobrevivência como segue:

```
Cálculo das razões de sobrevivência:
nLx[1:17,4] = nLx[2:18,3]/nLx[1:17,3]
nLx[18,4] = nLx[19,3]/(nLx[18,3]+nLx[19,3])
nLx[20:36,4] = nLx[21:37,3]/nLx[20:36,3]
nLx[37,4] = nLx[38,3]/(nLx[37,3]+nLx[38,3])
colnames(nLx) = c('sexo', 'idade', 'nLx', 'nSx')
```

A projeção da população fechada sem os nascimentos no quinquênio (tal como apresentada na coluna 5 da Tabela 21.6) é obtida como segue:

```
pop$nSx = nLx$nSx
pop[2:19,5] = round(pop[1:18,3]*pop[1:18,4],0)
pop[19,5] = round((pop[19,3]+pop[18,3])*pop[18,4],0)
pop[21:38,5] = round(pop[20:37,3]*pop[20:37,4],0)
pop[38,5] = round((pop[38,3]+pop[37,3])*pop[37,4],0)
colnames(pop) = c('sexo', 'idade', 'pop2010', 'nSx', 'pop2015fechada')
```

Uma vez projetada a população fechada, obtem-se a população aberta, já contabilizando os saldos migratórios por grupos etários e sexo, tal como apresentado na coluna 7 da Tabela 21.6:

```
pop$TLM = TLM$TLM
pop$pop2015aberta = round(pop$pop2015fechada+(pop$pop2015fechada*pop$TLM),0)
```

Uma vez projetada a população aberta, obtêm-se os nascimentos sobreviventes por sexo durante o quinquênio (2010-2015), ou seja, a população sobrevivente no grupo etário de 0-4 anos. Inicialmente, obtêm-se a contribuição das mulheres migrantes e não migrantes, nos grupos etários do intervalo reprodutivo (15-49 anos), para os nascimentos totais durante o quinquênio, cujo resultado é apresentado na coluna 9 da Tabela 21.6.

```
pop[4:10,8] = TEF$TEF
pop[1,9] = round(2.5*sum((pop[4:10,3]+pop[4:10,7])*pop[4:10,8]),0)
```

Em seguida, mediante o total de nascimentos obtidos anteriormente, separa-se as parcelas de nascimentos por sexo e determina-se os nascimentos sobreviventes ao longo do quinquênio (resultados apresentados na coluna 10 da Tabela 21.6):

```
pop[1,10] = round(pop[1,9]*(1/(1+1.04))*(nLx[1,3]/500000),0)
pop[20,10] = round(pop[1,9]*(1.04/(1+1.04))*(nLx[20,3]/500000),0)
colnames(pop) = c('sexo', 'idade', 'pop2010', 'nSx', 'pop2015fechada', 'TLM',
'pop2015aberta','TEF', 'B.total', 'B.sexo.sobrev')
```

Finalmente, obtêm-se a população projetada para 2015:

```
pop$pop2015 = round(ifelse(is.na(pop$pop2015aberta), pop$B.sexo.sobrev,
pop$pop2015aberta),0)
```

A planilha final com todas os cálculos e a população projetada para 2015 tem a estrutura mostrada na Tabela 21.6.

Tabela 21.6: Resultados do exercício de projeção do Estado do Rio Grande do Norte

	sexo	idade	pop2010	nSx	pop2015 fechada	TLM	pop2015 aberta	TEF	B.total	B.sexo.sobrev	pop2015
1	fem	0	126158	0.9983466	NA	0.00363	NA	NA	247728	119572	119572
2	fem	5	133256	0.9985941	125949	0.00455	126522	NA	NA	NA	126522
3	fem	10	145957	0.9984434	133069	0.01500	135065	NA	NA	NA	135065
4	fem	15	149374	0.9977187	145730	0.01039	147244	0.0660	NA	NA	147244
5	fem	20	156576	0.9967317	149033	0.00034	149084	0.0888	NA	NA	149084
6	fem	25	147147	0.9959023	156064	-0.03753	150207	0.0817	NA	NA	150207
7	fem	30	130679	0.9945915	146544	-0.00437	145904	0.0635	NA	NA	145904
8	fem	35	116755	0.9922171	129972	0.03438	134440	0.0329	NA	NA	134440
9	fem	40	112973	0.9880639	115846	0.04085	120578	0.0089	NA	NA	120578
10	fem	45	101144	0.9811114	111625	0.02134	114007	0.0008	NA	NA	114007
11	fem	50	78679	0.9722535	99234	0.03590	102797	NA	NA	NA	102797
12	fem	55	63016	0.9615464	76496	0.04756	80134	NA	NA	NA	80134
13	fem	60	57135	0.9442322	60593	0.06690	64647	NA	NA	NA	64647
14	fem	65	42481	0.9159959	53949	0.05402	56863	NA	NA	NA	56863
15	fem	70	35305	0.8658528	38912	0.06173	41314	NA	NA	NA	41314
16	fem	75	23538	0.7809689	30569	0.13123	34581	NA	NA	NA	34581
17	fem	80	18305	0.6649361	18382	0.30433	23976	NA	NA	NA	23976
18	fem	85	10685	0.4365564	12172	0.56211	19014	NA	NA	NA	19014
19	fem	90	6639	NA	7563	1.35873	17839	NA	NA	NA	17839
20	masc	0	131289	0.9974203	NA	0.00358	NA	NA	NA	125389	123566
21	masc	5	137770	0.9978030	130950	0.00449	131538	NA	NA	NA	131538
22	masc	10	151688	0.9945735	137467	0.01361	139338	NA	NA	NA	139338
23	masc	15	151873	0.9895731	150865	0.01031	152420	NA	NA	NA	152420
24	masc	20	157120	0.9869381	150289	-0.01005	148779	NA	NA	NA	148779
25	masc	25	144821	0.9862556	155068	-0.05353	146767	NA	NA	NA	146767
26	masc	30	125795	0.9834162	142831	-0.02015	139953	NA	NA	NA	139953
27	masc	35	109931	0.9788177	123709	0.03550	128101	NA	NA	NA	128101
28	masc	40	105020	0.9734926	107602	0.04573	112523	NA	NA	NA	112523
29	masc	45	92420	0.9636327	102236	0.02782	105080	NA	NA	NA	105080
30	masc	50	69628	0.9499493	89059	0.04084	92696	NA	NA	NA	92696
31	masc	55	53407	0.9306511	66143	0.05049	69483	NA	NA	NA	69483
32	masc	60	47903	0.9008398	49703	0.07002	53183	NA	NA	NA	53183
33	masc	65	34641	0.8623269	43153	0.08590	46860	NA	NA	NA	46860
34	masc	70	27776	0.8003049	29872	0.10704	33069	NA	NA	NA	33069
35	masc	75	17425	0.7018425	22229	0.15727	25725	NA	NA	NA	25725
36	masc	80	13370	0.5744033	12230	0.32553	16211	NA	NA	NA	16211
37	masc	85	7924	0.3709107	7680	0.62132	12452	NA	NA	NA	12452
38	masc	90	4336	NA	4547	1.28027	10368	NA	NA	NA	10368

Obtida a projeção da população do Estado do Rio Grande do Norte (Brasil) por sexo e grupos etários quinquenais para o ano de 2015, projeta-se a população total das microrregiões do Estado (área menores) pela aplicação do método Distribuição Constante do Crescimento, conforme descrito na seção 21.7.1.

Carregando os dados de entrada no software:

```
pop.menor = read.csv('pop.area.menor.csv', sep=";", stringsAsFactors = FALSE)
pop.maior = read.csv('pop.area.maior.csv', sep=";", stringsAsFactors = FALSE)
anos = pop.maior$ano
proj = pop.maior$popRN
```

Os arquivos “pop.area.menor.csv” e “pop.area.maior.csv” devem conter as seguintes estruturas:

pop.area.menor.csv:			pop.area.maior.csv:	
nome.micro	pop2000	pop2010	Ano	popRN
Mossoró	274976	332679	2011	3271415
Chap.Apodi	70290	72447	2012	3302720
Médio Oeste	37064	39041	2013	3332952
Vale do Açu	124753	140534	2014	3363084
Serra S.Miguel	58579	62755	2015	3393814
Pau dos Ferros	112349	114267		
Umarizal	64043	64984		
Macau	44836	52508		
Angicos	50932	51304		
Serra Santana	60682	61526		
Seridó Ocid.	89562	97680		
Seridó Oriental	114535	118828		
Baixa Verde	56667	62935		
Borborema RN	125767	134027		
Agreste RN	205710	229795		
Litoral NE	75188	84040		
Macaíba	240694	288836		
Natal	856579	1030764		
Litoral Sul RN	113576	129077		

Os objetos anos e proj devem ser vetores contendo, respectivamente, os anos calendários e as projeções da população total do Estado, tal como segue:

```
> anos
[1] 2011 2012 2013 2014 2015
> proj
[1] 3271415 3302720 3332952 3363084 3393814
```

Usando a variante (21.30) do método, a seguinte função produz os resultados desejados para cada microrregião nos anos calendários definidos no vetor “anos”:

```
DCC=function(mat,proj,anos){
  proj_armenor=matrix(0,dim(mat)[1],length(proj))
  for (i in 1:length(proj)){
    proj_armenor[,i]=mat[,3]+(proj[i]-sum(mat[,3]))*(mat[,3]-mat[,2])/
      (sum(mat[,3])-sum(mat[,2]))
  }
  colnames(proj_armenor)=t(anos[1:length(anos)])
  proj_armenor
}
```

Alternativamente, a seguinte função faz os cálculos necessários para resolver a variante (21.31), com o mesmo resultado, com os parâmetros adicionais a_i e b_i :

```
DCC=function(mat,proj,anos){
  ai=(mat[,3]-mat[,2])/(sum(mat[,3])-sum(mat[,2]))
  bi=mat[,2]-(ai*sum(mat[,2]))
  proj_armenor=matrix(0,dim(mat)[1],length(proj))
  for (i in 1:length(proj)){
    proj_armenor[,i]=ai*proj[i]+bi
  }
  resultado=cbind(ai,bi,proj_armenor)
  colnames(resultado)=cbind("ai","bi",t(anos[1:length(anos)]))
  resultado
}
```

Os argumentos das funções acima são:

mat = matriz ou dataframe de entrada (que deve ser o arquivo “pop.area.menor.csv”)

proj = vetor com a população total projetada tal como definido anteriormente.

anos = vetor com os anos calendários de projeção tal como definido anteriormente.

Uma vez carregada a função DCC descrita acima, para aplicá-la aos argumentos definidos acima, basta fazer:

```
proj.micros.RN = DCC(pop.menor,proj,anos)
```

onde “proj.micros.RN” é o nome do objetivo que receberá os resultados das projeções das populações totais das microrregiões. Para organizar a matriz final com os dados iniciais por microrregiões e seus respectivos cálculos, se faz:

`proj.micros.RN = cbind(pop.menor, proj.micros.RN)`

O objeto “proj.micros.RN” pode ser uma matriz ou dataframe que apresenta a seguinte estrutura com os resultados:

Tabela 21.7: Resultados do exercício de desagregação municipal da projeção do Estado do Rio Grande do Norte, usando o método de Distribuição Constante do Crescimento

	nome.micro	pop2000	pop2010	ai	bi	2011	2012	2013	2014	2015
1	Mossoró	274976	332679	0.14749	-134572	347929	352546	357005	361449	365982
2	Chap.Apodi	70290	72447	0.00551	54990	73015	73188	73355	73521	73690
3	Médio Oeste	37064	39041	0.00505	23041	39562	39720	39872	40025	40180
4	Vale do Açu	124753	140534	0.04034	12738	144707	145970	147189	148405	149644
5	Serra S.Miguel	58579	62755	0.01067	28951	63857	64191	64514	64835	65163
6	Pau dos Ferros	112349	114267	0.00490	98743	114773	114926	115074	115222	115373
7	Umarizal	64043	64984	0.00241	57351	65235	65311	65383	65456	65530
8	Macau	44836	52508	0.01961	-9617	54535	55149	55742	56333	56936
9	Angicos	50932	51304	0.00095	48294	51402	51432	51460	51489	51518
10	Serra Santana	60682	61526	0.00216	54684	61750	61818	61883	61948	62015
11	Seridó Ocid.	89562	97680	0.02075	31944	99826	100475	101103	101728	102366
12	Seridó Oriental	114535	118828	0.01097	84074	119961	120305	120636	120967	121304
13	Baixa Verde	56667	62935	0.01602	12183	64591	65093	65577	66060	66552
14	Borborema RN	125767	134027	0.02111	67149	136209	136869	137508	138144	138792
15	Agreste RN	205710	229795	0.06156	34771	236159	238086	239948	241802	243694
16	Litoral NE	75188	84040	0.02263	12349	86381	87090	87774	88456	89151
17	Macaíba	240694	288836	0.12305	-100989	301559	305411	309131	312838	316620
18	Natal	856579	1030764	0.44521	-379672	1076795	1090732	1104192	1117607	1131288
19	Litoral Sul RN	113576	129077	0.03962	3560	133173	134414	135612	136805	138023

Convém ainda assinalar que também existem vários programas de projeção demográfica em “R”, de domínio público e autoria de Eddie Hunsinger, na página web do Depto. de Demografia da Universidade de Berkeley (<https://applieddemogtoolbox.github.io/Toolbox/>).

21.9 ERROS, INCERTEZA E PROJEÇÕES PROBABILÍSTICAS

Segundo um aforismo famoso atribuído a Niels Bohr (mas também a Oscar Wilde, Mark Twain, Yogi Berra e à tradição oral dinamarquesa), “é muito difícil prever, especialmente o futuro”. Como qualquer tentativa de prever o futuro, uma projeção pode levar a uma previsão errônea. Erros de projeções não significam que a metodologia subjacente à projeção em si seja errada. Significam apenas que os supostos que orientaram a implementação do método não foram satisfeitos. Por exemplo, a projeção feita pela Divisão de População da ONU em 2015 para a Alemanha claramente não previu corretamente qual seria a população do país em 2020. O resultado publicado foi de 80,39 milhões de habitantes, mas isso não considerou o grande número de refugiados que o governo admitiu desde o segundo semestre de 2015. Na Revisão de 2017, foi necessário aumentar

a projeção para 82,54 milhões para dar conta dos acontecimentos. Na Revisão de 2019 aumentou outra vez, para 83,78 milhões. Nesse caso assumiu-se uma função de migrações que não correspondeu à realidade.

Keyfitz (1981) sistematizou os fatores que explicam a qualidade das previsões obtidas por meio de diferentes tipos de projeções. Embora o método de projeção tenha certa importância, o fator mais importante identificado por Keyfitz é o período em que uma determinada projeção foi preparada: em períodos sem grandes mudanças na dinâmica demográfica as projeções fornecem uma boa indicação do futuro, mas elas tipicamente não demonstram um bom desempenho na previsão de rupturas nas tendências. Keilman (2008) mostrou que – apesar das melhorias metodológicas introduzidas em décadas recentes – a qualidade das previsões demográficas na Europa não melhorou desde os anos 80, em boa parte devido a uma reversão não esperada nas tendências da fecundidade. Um bom exemplo é a análise que Carrilho (2005: Quadro 2) fez das projeções da fecundidade portuguesa formuladas em 1977. Estas previam um ligeiro aumento da fecundidade de mulheres com menos de 25 anos e um declínio suave da TFR em aproximadamente 15% até 1990. Mas o que efetivamente aconteceu foi um rompimento com a tendência histórica, em que a fecundidade caiu pela metade e em mais de 70% entre mulheres de mais de 35 anos. Algo parecido aconteceu no Brasil, onde a forte queda da fecundidade verificada a partir do Censo de 1980 (ver Gráfico 2.10) pegou a maioria dos demógrafos de surpresa. Guimarães (2014) analisa as dificuldades da previsão das tendências das componentes demográficas dentro de uma visão mais teórica baseada em estatística bayesiana.

Quanto maior o horizonte de tempo considerado maior é a incerteza em relação aos resultados obtidos numa projeção populacional, pois maior é o tempo de exposição à possibilidade de que eventos não esperados aconteçam. Após 30-40 anos, grande parte da população projetada será composta por pessoas que ainda não tinham nascido no momento da formulação da projeção, as quais podem ter padrões comportamentais ainda não descritos ou imaginados, o que aumenta a incerteza inerente aos resultados das projeções.

Os erros também podem acontecer por variações aleatórias em torno da função utilizada como pressuposto. Uma população que observa uma Taxa Bruta de Natalidade de 2 pessoas por mil habitantes em média, por exemplo, pode, em determinado ano, observar 3 nascimentos por mil habitantes simplesmente por uma variação aleatória em torno do valor estimado e observado em média durante todos os demais anos. Porém, o nascimento de algumas pessoas a mais ou a menos numa população pode se refletir em todas as demais funções da população por um efeito acumulativo em todos os demais anos, afetando a composição da população total. Assim, erros de projeção podem se dar pela definição incorreta dos pressupostos ou por variações aleatórias em torno dos pressupostos assumidos (Corrêa, 2014; Olivieri e Pitacco, 2011). Assim, considerando conjuntamente a incerteza proveniente do tempo de projeção e da aleatoriedade dos eventos demográficos, o recomendado é que para populações pequenas o horizonte de tempo analisado seja menos que para populações maiores e menos afetadas por variações aleatórias.

Assim, a projeção será mais precisa para áreas geográficas com populações maiores do que para aquelas com populações menores (Hoque, 2010). Por exemplo, as estimativas populacionais tendem a ser mais precisas para um país inteiro ou estado em relação a subáreas dentro do país ou estado. Desastres naturais, fechamentos ou instalações de fábricas, entre outros, podem ter um impacto importante sobre a população na área imediata, mas não terá praticamente nenhum

impacto sobre a população nacional. A projeção também será mais precisa para as populações totais e não para subgrupos da população (Hoque, 2010), como os definidos por idade, sexo, raça e etnia, situação socioeconômica, pois, quanto maior o nível dos detalhes, maior a possibilidade de erro. A precisão também será maior se utilizados dados que determinam diretamente a mudança da população (tais como dados sobre nascimentos, mortes e migração) do que com os dados que empregam indicadores indiretos ou sintomáticos das alterações demográficas (como o recenseamento eleitoral, registro de veículos e matrículas escolares) (Hoque, 2010).

Em função da incerteza quanto à realização dos resultados projetados, em métodos que projetam um único resultado possível para a projeção, os chamados métodos determinísticos, é comum adotar-se mais de um cenário de projeção, como um cenário com níveis de fecundidade alto, médio e baixo, por exemplo, ou para as outras funções demográficas. Entretanto, ao assumir cenários não se define a probabilidade de ocorrência de cada um desses cenários estimados, nem se define a possibilidade de resultados que possam ocorrer entre dois cenários definidos.

Diante disso existem vários métodos de projeções populacionais que estimam um valor médio e intervalos de confiança em torno desse valor médio, incorporando as possibilidades de erro de estimação pela aleatoriedade das funções utilizadas e a incerteza inerente do tempo pelos chamados métodos probabilísticos. Entre esses, algumas metodologias se voltam à projeção de funções demográficas em si; considerando a variabilidade dessas funções em seus cálculos (Lee e Carter, 1992) desenvolveram o método mais usual para a projeção da função de mortalidade que gera um intervalo de confiança probabilístico para as projeções (ver seção 21.2). O método utiliza séries temporais do tipo ARIMA para projetar a mortalidade total e por grupo etário, mantendo a estrutura etária de mortalidade constante. Contudo, a confiança desse método de séries temporais está relacionada à abundância de uma série histórica de dados de mortalidade, o que nem sempre é possível, principalmente a nível local. O modelo também tem sido generalizado para outros componentes da dinâmica demográfica (fecundidade e migração), resultando assim numa metodologia probabilística completa de projeção demográfica que tem sido aplicado para projetar a população da Grã-Bretanha para 2024 (Wisniowski et al., 2015).

Métodos de estimação e projeção de fecundidade também introduziram medidas de variabilidade nas projeções. Doblhammer, Milewski e Peters (2011) propõem um método de estimação de fecundidade mensal com informações de nascimentos mês a mês e estima, inclusive, o intervalo de confiança para os eventos. Já Alkema et al. (2011) propõem um método probabilístico de projeção da fecundidade baseado em 3 fases de mudança da fecundidade durante a transição demográfica: a primeira com altas Taxas de Fecundidade Totais (TFT), a segunda com a queda acentuada das TFT, e a terceira em que as TFT abaixo no nível de reposição aumentam gradualmente até a marca de 2,1, considerada como nível de reposição. Para cada país uma trajetória diferente é projetada com base nos dados históricos do próprio país por meio de simulações de Monte Carlo usando cadeias de Markov, as quais geram, também, intervalos de confiança para a TFT. Posteriormente, a mesma metodologia foi generalizada para outros componentes da dinâmica demográfica (mortalidade e migração) e usada para fazer projeções probabilísticas para todos os países do mundo (Raftery et al., 2012). Atualmente, esta é a metodologia em que se baseiam as projeções da Divisão de População das Nações Unidas.

Também Abel et al. (2013) usam simulações de Monte Carlo usando cadeias de Markov, mas para estimar vários modelos de séries temporais (modelo autorregressivo, de volatilidade

estocástica, de mudança de variância aleatória) com variância constante e não constante para projeções populacionais com intervalos de confiança. Os mesmos métodos poderiam ser utilizados para as funções demográficas de fecundidade, mortalidade, migração, ou outra função demográfica.

Nesses métodos, tal como no Método de Lee e Carter para a mortalidade, o objetivo é a estimação de uma função demográfica média. A variância estimada por esses métodos, portanto, se refere à estimação do intervalo de confiança da função média, e não à estimação da variância da função demográfica em si, tanto que não guardam relação com o tamanho populacional.

Já os métodos de microsimulação (ver seção 17.2 do Capítulo 17) conseguem relacionar a aleatoriedade das funções demográficas com o tamanho populacional. Por se basear em sorteios aleatórios, a cada simulação o resultado obtido pode ser diferente de forma que, ao se gerar uma grande quantidade de resultados diferentes, pelo Método de Monte Carlo, pode-se obter uma gama de resultados e uma distribuição de probabilidade dos resultados possíveis. Os métodos de microsimulação são mais complexos computacionalmente, mas os avanços computacionais permitem que sejam mais utilizados a cada dia.

21.10 ESTIMATIVAS BASEADAS NO USO DE VARIÁVEIS SINTOMÁTICAS

Embora as estimativas de população baseadas em variáveis sintomáticas estritamente falando não sejam projeções, o tema geralmente é tratado nesse contexto. A ideia básica é o uso de dados de registros administrativos (variáveis sintomáticas), como o número de nascimentos, de mortes, de novas unidades de moradia, de matrículas escolares, consumo de energia e água, juntamente com dados correlatos do último censo, para estimar a população total de uma pequena área (Rao, 2003). Nesse sentido, o método é parecido com os métodos estruturais que foram mencionados na introdução deste capítulo, mas com a diferença de que os indicadores usados são efetivamente observados e não projetados. Mais especificamente, às vezes se usam as estatísticas locais de nascimentos e óbitos para estimar a população (Bogue, 1950). Existem diversos variantes do método: na sua revisão de métodos para a projeção de pequenas áreas, González e Torres (2012) identificam nada menos de nove. O suposto geral da metodologia é que deve haver uma certa proporcionalidade entre todas essas variáveis sintomáticas e a população que as produz e que essa proporcionalidade se mantém ao longo do tempo. Ou seja, se o número de matrículas escolares no ano do último censo foi $x\%$ do número de habitantes, supõe-se que a mesma relação deve existir na atualidade, o que permite estimar a população atual a partir do número de matrículas registradas.

Entretanto, para nenhuma das variáveis consideradas a proporcionalidade é perfeitamente constante. O número de matrículas pode variar em função de mudanças na composição da população (mais ou menos crianças) ou de mudanças na proporção de crianças de uma determinada idade que se matriculam. Da mesma forma, o número de óbitos é sensível à estrutura etária, especificamente à proporção de pessoas idosas. O consumo de energia, por sua vez, não depende só do número de consumidores potenciais, mas também do preço da energia, de mudanças no poder aquisitivo da população e de fatores conjunturais como o tempo (por exemplo, mais uso de ar condicionado em épocas de calor). Portanto, em vez de usar uma só variável se usam várias, na esperança de que os erros se compensem.

Uma das variantes mais comuns do método é aquela onde se dispõe de k variáveis para descrever a população de n unidades menores num ano t em que se conhece também a população destas

unidades. Sendo assim, pode ser formulada uma equação de regressão que estima a população como função linear das variáveis sintomáticas:

$$P_i(t) = \alpha_0 + \alpha_1 VS_{I_i}(t) + \dots + \alpha_k VS_{k_i}(t) + Erro(i,t) \quad (21.51)$$

Geralmente impõe-se a condição de que $\alpha_1 = 0$. Agora, aplicando-se os mesmos coeficientes em $t+\Delta t$ é possível estimar $P(t+\Delta t)$. Se as n unidades formam parte de uma área maior para a qual existe uma projeção em $t+\Delta t$, é preferível usar a seguinte fórmula proporcional:

$$P_i(t) / P(t) = \alpha_0 + \alpha_1 VS_{I_i}(t)/VS_{I_1}(t) + \dots + \alpha_k VS_{k_i}(t)/VS_{k_k}(t) + Erro(i,t) \quad (21.52)$$

onde $VS_{I_i} \dots, VS_{k_i}$ se referem às variáveis sintomáticas da unidade i e $VS_{I_p} \dots, VS_{k_k}$ aos totais da área maior.

Tanto (21.51) como (21.52) sofrem do problema de colinearidade entre as diferentes variáveis sintomáticas o qual pode fazer com que os coeficientes de regressão sejam bastante instáveis. Por isso pode ser preferível usar uma média simples ou uma mediana dos resultados univariados, em vez de calcular coeficientes multivariados de regressão. Ou seja, se calcula

$$\alpha_1 = \frac{P_i(t)}{VS_{I_i}(t)} \quad \text{ou} \quad \alpha_1 = \frac{P_i(t)/P(t)}{VS_{I_i}(t)/VS(t)} \quad (21.53)$$

dependendo da especificação do modelo: (21.51) ou (21.52). Em seguida, se calcula a média simples ou a mediana de $\alpha_p, \dots, \alpha_k$.

Em vez de estimar a população total, uma alternativa é estimar diferentes subpopulações, considerando que o número de nascimentos é mais expressivo da população de 0-4 anos, o número de matrículas tem mais relação com a população de 5-14 anos e o número de óbitos mantém uma relação mais estreita com a população de mais de 60 anos.

Existem várias aplicações de diferentes variantes da metodologia ao Brasil e América Latina. Alguns exemplos incluem Jardim (1985), Freire (1997), Bay (1998), González (2010) e Barros e Cavenaghi (2016).