

MÉTODOS DE ESTIMAÇÃO INDIRETA DA MORTALIDADE E FECUNDIDADE

23.1 INTRODUÇÃO

A obtenção de estimativas demográficas confiáveis do nível e padrão da mortalidade e da fecundidade é muito importante para a compreensão da dinâmica demográfica, no planejamento fiscal e de políticas públicas e sociais, especialmente nas áreas discutidas no Capítulo 3. Em países com bons registros vitais, estas estatísticas podem ser calculadas com os métodos descritos nos Capítulos 8 e 10. No caso dos países em desenvolvimento, no entanto, estimar essas medidas se torna um desafio, uma vez que a qualidade da informação dos registros vitais geralmente não é satisfatória. As dificuldades existentes nas fontes de dados utilizadas na estimativa da mortalidade estão relacionadas à incompleta cobertura do registro de óbitos e a erros na declaração da idade, tanto no registro de óbito quanto nos dados de população (Hill, 2017; Hill, You e Choi, 2009, entre outros). Os erros na declaração de idade são menos importantes na estimativa da fecundidade, mas os registros de nascimentos sofrem, além de erros de sub-registro definitivo, do registro atrasado (ver Capítulo 5). Até certo ponto, é possível obter informação de melhor qualidade por meio de inquéritos do tipo DHS ou MICS, mas estes nem sempre estão disponíveis e as suas características amostrais limitam o seu uso em nível local.

Para contornar essas limitações dos dados, diferentes metodologias foram desenvolvidas para medir a mortalidade e a fecundidade de forma *indireta* (Moultrie et al., 2013). O termo “indireto” se refere ao fato de as estimativas não serem obtidas diretamente como seria o caso ao usar fontes convencionais (o registro civil ou vital) e de precisar, com frequência, assumir pressupostos baseados nos conceitos de populações teóricas. Um dos primeiros trabalhos que aplicou estes métodos

aos PALOP foi o de Heisel (1968). Hoje este trabalho está muito desatualizado, mas tanto o texto em si como a coletânea de textos sobre a demografia africana da qual faz parte fornecem uma visão interessante desde o ponto de vista da história dos métodos. Os elementos usados mais frequentemente como alternativa ou correção do registro civil ou vital são os seguintes:

1. A estrutura da população por idade e sexo estabelecida pelo censo ou inquérito amostral;
2. A distribuição por idade, sexo e parentesco dos menores de 15 anos e coenumeradas no censo, por idade das mulheres que supostamente são as suas mães;
3. Perguntas retrospectivas feitas no censo ou em inquéritos amostrais que fornecem informações sobre os números totais de nascimentos de filhos das mulheres enumeradas e sobre a sobrevivência desses filhos e/ou outras pessoas (pais, mães, irmãos e irmãs) de interesse;
4. No censo normalmente não se pergunta quais foram as datas de nascimento de todos os filhos nascidos vivos (exceto, eventualmente, o último), mas em inquéritos do tipo DHS ou MICS se levanta uma história de nascimentos mais completa que inclui esses dados;
5. A pergunta sobre óbitos ocorridos no domicílio (agregado familiar) durante certo período (tipicamente, um ano) no passado recente; e
6. Os modelos teóricos de mortalidade e fecundidade discutidos no Capítulo 20 que permitem inferir certas relações plausíveis entre os indicadores.

Além disso, alguns dos métodos usam dados dos próprios registros, no intuito de corrigi-los. Naqueles casos onde existem dados do tipo 4), os procedimentos são muito parecidos com a forma de analisar informação do registro civil ou vital. A diferença principal é que os dados são captados retrospectivamente, de modo que excluem, por exemplo, informação sobre os filhos de mulheres que já morreram ou que migraram. As histórias de nascimentos, à diferença das estatísticas vitais, permitem estimar indicadores adicionais de fecundidade, como, por exemplo, o intervalo entre nascimentos. Os dados do tipo 2) são usados para uma técnica muito específica, chamada o “método dos filhos próprios” que permite a estimação da fecundidade, mas não da mortalidade e que será discutida na seção 23.3.3 deste capítulo. Teoricamente, seria possível combinar os outros tipos de dados (números 1, 3 e 5) com os modelos em 6) para desenvolver um único método integrado para a estimação simultânea da fecundidade e mortalidade, mas na prática isso é difícil devido às variações no formato dos dados disponíveis e na confiabilidade da informação para diferentes propósitos. Por exemplo, dados do tipo 5) são muito úteis para a estimação da mortalidade adulta, mas não para a estimação da mortalidade infantojuvenil porque os óbitos de crianças muito jovens muitas vezes são omitidos. Por isso geralmente são usadas diferentes partes da informação disponível para cada propósito. Para analisar a fecundidade e a mortalidade infantojuvenil geralmente se usam 3) e 6); para estimar a mortalidade adulta se usam 1), 5) e 6).

Este capítulo discutirá primeiro os métodos para a estimação (ou correção) da fecundidade que tendem a ser mais simples porque na maioria dos casos não exigem o conhecimento da mortalidade, desde que não haja uma relação sistemática entre ambas. O segundo assunto, tratado

na seção 23.4, é a estimação da mortalidade infantojuvenil com base em perguntas retrospectivas sobre filhos sobreviventes. Os resultados desses métodos dependem não só da proporção de filhos sobreviventes, mas também do padrão de fecundidade, na medida em que este determina há quanto tempo, na média, esses filhos nasceram. Finalmente, a seção 23.5 trata da estimação da mortalidade adulta.

Outra vez é preciso alertar para o fato de que os métodos expostos neste capítulo precisam ser usados com critério e não de forma mecânica. Há uma variedade de métodos que precisam ser bem ajustados aos dados existentes e todos os cuidados devem ser tomados na análise. Antes de aplicá-los, sempre se deve avaliar a qualidade dos dados e os vieses prováveis que os afetam, inclusive os diferenciais entre diferentes grupos e as tendências temporais. A escolha do método de análise deve tomar esses fatores em conta. Muitos dos métodos possuem supostos fortes e se esses não forem satisfeitos os resultados podem ser absurdos. Por exemplo, se o método P/F (ver abaixo) indica a necessidade de fazer ajustes de mais de 100% para cima ou 50% para baixo é bom se questionar se os supostos do método realmente se aplicam aos dados analisados e se há necessidade de uma reinterpretação dos resultados. Sempre que possível, é recomendável comparar diferentes alternativas com diferentes dados. Isso ajuda a obter resultados mais robustos.

23.2 DADOS USADOS NA APLICAÇÃO DOS MÉTODOS INDIRETOS

Alguns dos elementos na lista da seção anterior se referem a dados que de todos os modos já são recolhidos nos censos e outros instrumentos, como os elementos 1) e 2). O 6) não é um dado, mas um instrumento metodológico. Mas os elementos 3), 4) e 5) exigem informações específicas que precisam ser levantadas nos censos ou em inquéritos. Portanto, é preciso fazer uma breve revisão da informação disponível sobre o tema nas diferentes fontes.

No caso da fecundidade, as perguntas específicas feitas no censo são basicamente as seguintes:

- a) Número de filhos nascidos vivos tidos pelas mulheres, independentemente da sobrevivência destes (de preferência desagregados por sexo). O critério de idade para inquirir a mulher varia segundo a necessidade do país. Pode ser 10-50 anos (Moçambique em 2017) ou todas as mulheres maiores desde 12 ou 15 anos. O mais comum é considerar as idades de 15-49 anos completos. O dado obtido serve para medir a fecundidade retrospectiva. Alguns países limitam a pergunta a mulheres não solteiras. Esta pergunta é feita nos censos de quase todos os países, inclusive aqueles que têm boas estatísticas vitais. Portugal a incluía no seu censo até 1991, mas a retirou em 2001 e 2011. Alguns dos censos portugueses inclusive distinguem os filhos do último casamento dos filhos de casamentos anteriores. A informação geralmente é publicada por grupos etários convencionais, de modo que o cálculo das parturições médias P_1 , P_2 , P_3 etc. é direto.
- b) Os censos brasileiros estão entre os poucos que também perguntam sobre filhos nascidos mortos, embora essa informação não tenha um papel direto nos métodos de estimação discutidos aqui. A pergunta foi incluída pela primeira vez em 1940, retirada em 1950 e 1960, mas desde 1970 faz parte de todos os censos, sendo cortada outra vez em 2021. O Censo da Guiné-Bissau, de 2009, não perguntou diretamente sobre o número de natimortos (nados

mortos), mas perguntou o número de partos, além do número de nascidos vivos. Os censos coloniais de Moçambique, que desde 1940 perguntaram sobre o número de filhos tidos, geralmente se limitaram a filhos nascidos vivos, mas o Censo de 1950 também incluiu os natimortos (nados mortos). Além de ser uma informação sobre mortalidade intrauterina, é útil na avaliação da qualidade/confiabilidade da informação sobre nascidos vivos, uma vez que haveria uma proporção esperada entre ambos.

- c) Número de filhos nascidos vivos (independentemente da sua sobrevivência) nos 12 meses antes do censo ou data de nascimento do último filho. No Brasil esta pergunta foi introduzida pela primeira vez no Censo de 1970. Muitos países também distinguem, tanto aqui como na pergunta a), entre nascimentos de meninos e meninas.
- d) Poucos países (África do Sul, Namíbia, Catar, Laos e alguns países do Caribe) também perguntam a data de nascimento do primeiro filho ou filha ou a idade que a mulher tinha naquele momento. Esta pergunta não contribui muito para o conhecimento da fecundidade, mas é potencialmente útil para quantificar o tempo de exposição dos filhos à mortalidade. Isso exige técnicas de análise específicas. Outra possibilidade é o uso da informação sobre a data do primeiro casamento ou a idade que a mulher tinha naquele momento. O número de países que incluem esta pergunta no seu censo é um pouco maior, mas não é o caso de nenhum país de língua portuguesa.
- e) Alguns países, como Timor-Leste em 2015, também perguntam onde o parto ocorreu e que tipo de assistência médica a mãe recebeu. Esta informação não é necessária para estimar a fecundidade. No entanto, uma vez que se pergunta pelo nascimento dos filhos, é conveniente formulá-la junto com as perguntas sobre fecundidade quando se requer este tipo de informação.

Como foi mencionado anteriormente, a informação recolhida nas pesquisas do tipo DHS ou MICS é mais completa na medida em que esses inquéritos contêm histórias de vida de todos os filhos nascidos vivos, incluindo tanto as datas de nascimento como as datas de falecimento, onde for o caso.

A variedade da informação sobre a mortalidade é maior do que no caso da fecundidade. Aqui se podem distinguir as seguintes perguntas:

- a) Entre os filhos nascidos vivos declarados em a) (de preferência de cada sexo), quantos estavam vivos na data do censo ? Esta pergunta é feita em todos os censos dos países de língua portuguesa menos Portugal e Macau. No Brasil ela consta do Censo de 1890, mas depois foi retirada e reintroduzida, por sugestão do Giorgio Mortara, no Censo de 1940. Nas ex-colônias portuguesas, como Moçambique, a pergunta foi introduzida em 1940. Se a informação está disponível por sexo dos filhos, além de permitir o estudo da mortalidade segundo esta variável, é possível, também, avaliar a qualidade/confiabilidade da informação uma vez que espera-se, salvo raras exceções, uma sobremortalidade masculina.

- b) Alguns países, como Angola em 2014 e Moçambique em 2007 e 2017, também perguntam quantos dos filhos nascidos nos últimos 12 meses declarados em c) ainda estavam vivos na data do censo. Quando a pergunta c) é formulada em termos da data de nascimento do último filho (ou filha), frequentemente verifica-se a sua sobrevivência.
- c) Desde 1980, o censo brasileiro pergunta se a mãe da pessoa ainda está viva. Angola também fez esta pergunta em 2014. Além disso geralmente se pergunta, em caso afirmativo, se a mãe faz parte do mesmo domicílio (agregado familiar). Alguns censos, como o de Cabo Verde de 2010, Moçambique de 2007 e 2017 e Timor-Leste de 2015, também perguntaram sobre a sobrevivência do pai, mas esta informação é considerada menos confiável. Embora atualmente a finalidade principal da pergunta sobre orfandade seja a estimação da mortalidade adulta, podem existir outras razões para incluí-la nos censos. Em alguns censos africanos foi feita para captar a *geração faltante* (“skip generation”, em inglês) devastada pela epidemia de HIV/AIDS (sobre este assunto, ver Richter e Desmond, 2008). Em Portugal, a pergunta tradicionalmente era feita, muito antes que começasse a ser aproveitada para estimar a mortalidade adulta, mas foi retirada em 1981.
- d) Outra forma para medir a mortalidade é perguntando quantas pessoas faleceram no domicílio (agregado familiar) nos últimos 12 meses, por sexo, idade e data do falecimento (alguns países omitem esta informação). O propósito destas perguntas é compor um registro de óbitos paralelo ao do registro civil para o ano anterior à data do censo. No Brasil, este procedimento foi usado nos Censos de 1980, 2010 e 2020. Em 1980, a cobertura de óbitos foi apenas 40%, o que dificultou o uso desta informação, mas em 2010 a qualidade de informação foi consideravelmente melhor, com uma cobertura estimada de 80% (Queiroz e Sawyer, 2012). As perguntas também foram feitas no Censo de Angola de 2014, no Censo da Guiné-Bissau de 2009, nos Censos de Moçambique de 2007 e 2017 e nos Censos de Timor-Leste de 2010 e 2015.
- e) Nos PALOP que fizeram as perguntas descritas em d) acrescentou-se uma pergunta mais (de preferência desdobrada em três componentes) na rodada (ronda) dos censos de 2010, no caso de tratar-se de uma mulher em idade fértil. Ao perguntar se a mulher estava grávida, estava dando à luz ou tinha dado à luz nas últimas 6 semanas, cria-se uma oportunidade para estimar a mortalidade materna (ver seção 8.9 do Capítulo 8).
- f) Outra pergunta que pode ser usada para estimar a mortalidade materna investiga a sobrevivência das irmãs da pessoa entrevistada e sua situação de gravidez (mesmas categorias de e) no momento da sua morte. De uma forma geral a pergunta tem sido feita nos inquéritos do tipo DHS ou MICS devido à sua especificidade, apesar das amostras relativamente pequenas destes inquéritos. A lógica do método é parecida com a estimação da mortalidade infantojuvenil que será discutida na seção 23.4, mas os detalhes do cálculo não serão analisados aqui. Para uma discussão detalhada do método, o leitor é referido a WHO/UNICEF (1997).

23.3 MÉTODOS DE CORREÇÃO DA FECUNDIDADE

23.3.1 Método de El Badry

Um dos primeiros problemas que costuma apresentar-se é um número relativamente alto de mulheres com parturição desconhecida, principalmente em mulheres jovens. Cavenaghi (2015) cita o caso de Bolívia e Costa Rica na rodada de censos de 2010, onde mais de 20% das mulheres menores de 20 anos não declararam a informação sobre filhos tidos. Nestes casos, é aconselhável, primeiro, uma avaliação e, se possível, correção dos dados com base no método de *El-Badry*. Este ajusta os erros no número de nascidos vivos, causados por falha do recenseador ao deixar em branco a resposta sobre o número de filhos tidos. Durante o processamento dos dados, a resposta é codificada como “perdida” ou “desconhecida”, embora fosse evidente para o recenseador no momento de registrar o dado que a resposta era zero filho. Antes de aplicar o método descrito abaixo, geralmente se aplica algum algoritmo em nível individual, para ver se certos valores desconhecidos podem ser atribuídos. Por exemplo, se uma mulher solteira de 17 anos que mora com os pais, num grupo doméstico sem crianças menores de 5 anos, e frequenta a escola, tiver um número de filhos desconhecido, é altamente provável que o número real seja zero. Mas em outros casos, pode ser mais difícil fazer estas determinações.

A mecânica do método é bastante direta e simples. Reparte-se o número de mulheres cuja parturição é registrada como “perdida” entre: 1) aquelas cuja parturição é considerada como realmente desconhecida; e 2) as mulheres que deveriam ter sido registradas como sem filhos, mas cujas respostas foram deixadas em branco. É importante destacar que o método faz esta separação no nível agregado e não individualmente.

Para sua aplicação El-Badry exige conhecer o número de crianças nascidas vivas, classificado pela idade da mãe, incluindo a contagem de mulheres sem informação ou em branco, ou contém um código para não resposta ou recusa (por exemplo, 99 ou 98). Como pressuposto, ele assume que uma proporção constante de mulheres em cada idade realmente não indica sua parturição, no momento da coleta (recolha) de dados. As outras mulheres com parturição não declarada seriam erroneamente registradas como tendo informação faltante, quando estas mulheres, de fato, não tiveram filhos. Em termos práticos, o método se baseia na existência de uma *relação linear* entre a proporção de mulheres cuja parturição não é indicada e as mulheres que relataram não ter filhos.

O procedimento exige alguns cuidados:

- Se uma relação linear é observada entre essas medidas, o denominador usado para calcular parturições médias, deve excluir aquelas mulheres cuja parturição (após a correção) ainda é considerada desconhecida. Isso supõe que a distribuição de parturição dessas mulheres não é diferente daqueles das mulheres da mesma idade cuja P_i é conhecida.
- Caso os dados indiquem que a correção é necessária devido à grande proporção de falta de informação de parturição, mas o método não pode ser aplicado (por exemplo, devido à indisponibilidade de dados por idade, ou a violação do suposto de linearidade), as mulheres de parturição desconhecida devem ser incluídas no denominador usado para determinar as parturições médias.

- Assume-se que a parturição de todas essas mulheres é zero; ou seja, que todas as mulheres com parturição desconhecida não tiveram filhos. Isto, claro, resulta em parturições médias subestimadas, uma vez que nem todas as mulheres de parturição desconhecida realmente não tiveram filhos.

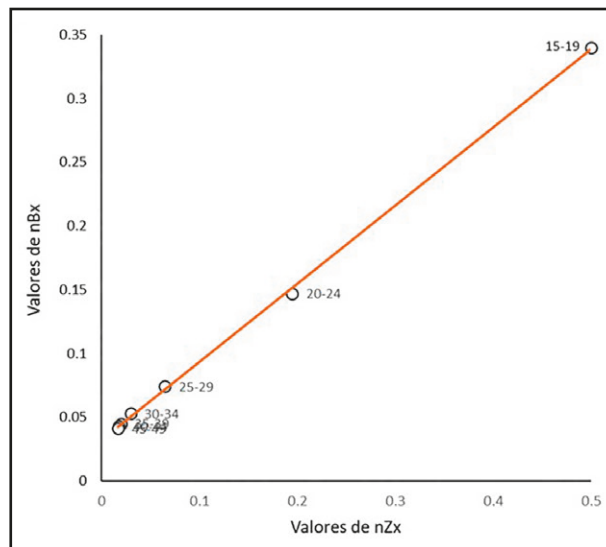
Em termos práticos, estima-se ${}_nB_x$ (proporção de mulheres de idade $(x, x+n)$ cuja parturição é desconhecida, ou seja, preenchida em branco) e ${}_nZ_x$ (proporção de mulheres com parturição zero).

$${}_nB_x = \frac{{}_nP_{x,d}}{{}_nP_x} \text{ e } {}_nZ_x = \frac{{}_nP_{x,z}}{{}_nP_x} \quad (23.1)$$

onde ${}_nP_{x,d}$ é o número de mulheres em idade x de cuja parturição é desconhecida; ${}_nP_{x,z}$ é o número de mulheres em idade $(x, x+n)$ de cuja parturição é zero. ${}_nP_x$ é o número de mulheres com idade entre x e $x+n$. Caso D seja pequeno (menos de 2%), não vale a pena aplicar o método. Assume-se que a distribuição de parturição desconhecida é idêntica à das mulheres cuja parturição é conhecida, de modo que essas podem ser eliminadas do denominador no cálculo dos ${}_nP_x$.

O próximo passo envolve uma análise gráfica dessas duas proporções, tal como apresentado no Gráfico 23.1. O conjunto de pontos deve formar uma reta. Pode ocorrer que alguns pontos não se alinhem num formato de reta (geralmente nas idades mais avançadas e jovens). Se esta curvatura afeta apenas as idades mais velhas, é possível excluir o último (ou dois últimos) grupos.

Gráfico 23.1: Relação entre a proporção de mulheres com número de filhos desconhecido (${}_nB_x$) e a proporção de mulheres sem filhos (${}_nZ_x$), Quênia 1989



Fonte: Moultrie et al., 2013: Figure 4.1.

Caso os grupos etários, que não se ajustam, sejam os das idades mais jovens, não é aconselhável excluir esses pontos. Além disso, caso a relação entre as duas proporções não for linear, não se deve aplicar o método, e se assume que todas as mulheres com informação desconhecida de parturição não tiveram filhos. Porém, elas entram no denominador do cálculo das parturições médias.

Uma vez verificada uma relação linear, deve-se determinar uma regressão linear com inclinação (B) e intercepto (A). Uma regressão linear entre ${}_n B_x$ em função de ${}_n Z_x$. Assim, ${}_n B_x = A + B {}_n Z_x$. O intercepto A estima a proporção de mulheres em cada idade com parturição realmente desconhecida. Valores muito altos de A indicam má qualidade dos dados. Assim, a proporção ajustada das mulheres na faixa etária $(x, x+n)$ que efetivamente não tiveram filhos é dada por: ${}_n Z'_x = {}_n Z_x + {}_n B_x - A$. Noutras palavras, a proporção revisada de mulheres com parturição zero, em qualquer faixa etária, é igual à proporção de mulheres com zero parturição mais a proporção de mulheres nessa faixa etária de parturição não declarada, menos a proporção estimada de mulheres cuja parturição é considerada como *verdadeiramente desconhecida*.

A estimativa revisada do número de mulheres sem filhos na faixa etária $(x, x+n)$ é dada por: ${}_n P'_{x,z} = {}_n P_x {}_n Z'_x$ e o número de mulheres com ${}_n P_x$ realmente desconhecido é dado pela proporção verdadeira de mulheres com parturição desconhecida: ${}_n P'_{x,d} = {}_n P_x A$. Um exemplo ilustrado da aplicação deste método, em que se baseia o Gráfico 23.1, pode ser visto em Moultrie et al. (2013).

23.3.2 P/F de Brass e suas variantes

O método P/F de Brass é um dos métodos mais usados para estimar o nível da fecundidade corrente e avaliar a qualidade dos dados provenientes da declaração das mulheres do domicílio, em idade reprodutiva, do número nascidos vivos, geralmente nos 12 meses anteriores à data de referência do censo ou pesquisa. Foi desenvolvido por Brass, em 1964 e está descrito no escrito no *Manual X* (Naciones Unidas, 1986), em conjunto com uma série de variantes usadas para estender o método, dependendo dos dados disponíveis. Um refinamento mais recente do método P/F é o modelo relacional de Gompertz (Naciones Unidas, 1986; Moultrie et al., 2013; Lima, Queiroz e Zeman, 2018) que será exposto na seção 23.3.2.3.

Como explicado no Capítulo 10, as informações que permitem corrigir os dados de fecundidade corrente, tipicamente obtidas no censo, ou pesquisa, são a parturição (P) feminina, uma medida de coorte, e a função acumulada de fecundidade corrente (F), uma medida reprodutiva de período. Assim, Brass (1968, 1975) propõe a razão P/F , em uma ou duas idades, dentro do período reprodutivo que se baseia na comparação entre a fecundidade de uma única coorte e a fecundidade corrente das mulheres acumulada até a idade dessa coorte (cruzamento de informações de várias coortes). Ao calcular a razão desses dois indicadores, tem-se um fator de ajuste para corrigir o erro de período de referência na declaração de fecundidade corrente. Em termos formais, existe a seguinte relação:

$$p(x, t) = \int_x^x f(y, t + y - x) dy \quad (23.2)$$

onde $p(x, t)$ é a parturição média de mulheres de idade exata x , y é uma idade menor ou igual a x (ou seja, $y \leq x$) e $f(y, t)$ é a fecundidade real de mulheres de idade exata y no momento t . Ambas são consideradas e notadas como funções contínuas. Na prática, o que se tem não são as parturições $p(x, t)$, mas as suas médias sobre os intervalos de 15-19, 20-24, 25-29 anos etc. Em termos formais, isso significa

$$P_i(t) = \int_{x_i}^{x_i+n} P^f(x, t) \int_{-\infty}^x f(y, t + y - x) dy dx / \int_{x_i}^{x_i+n} P^f(x, t) dx \quad (23.3)$$

onde $x_1=15$, $x_2=20$, $x_3=25$ etc., $n=5$ e $P^f(x, t)$ é a função de densidade da população de mulheres para a idade exata x e o momento t . Se (23.3) for aplicado a dados empíricos, contém os seguintes supostos implícitos:

1. As mulheres que morreram, que, portanto, não foram entrevistadas, percorreram a mesma função $f(y, t)$ que as mulheres sobreviventes;
2. As mulheres em cada grupo quinquenal de idade, dentro do período reprodutivo, estão corretamente classificadas;
3. Não há erros de seletividade sobre o total de filhos tidos nascidos vivos (a tendência de omitir filhos que já morreram) nas respostas das mulheres até as idades 30 ou 35 anos de idade;
4. Não existe diferencial de fecundidade entre mulheres migrantes e não migrantes até os 30 ou 35 anos de idade;
5. O erro de período de referência não varia com a idade das mulheres.

Todas as funções em (23.3) são as funções verdadeiras, algumas das quais são desconhecidas. O objetivo final é a estimação de $f(y, t)$, a fecundidade real no momento t , mas é preciso considerar que os dados observados não correspondem exatamente às funções em (23.3). Especificamente:

1. Observa-se a sequência empírica f_p, \dots, f_7 da fecundidade dos últimos 12 meses para as faixas etárias 15-19, ... , 45-49. Supõe-se que o *padrão etário* desta sequência é correto, mas que o seu *nível* sofre de erros de período de referência;
2. Se for obtida com base no censo, a fecundidade dos últimos 12 meses corresponde a grupos etários de mulheres defasados em meio ano, pois as mulheres de 20-24 anos (20 até 25, exclusive) tinham, na média, entre 19,5 e 24,5 anos quando os seus filhos nasceram (defasagem $\Delta=0,5$). Se os dados sobre a fecundidade recente provêm do registro civil, essa defasagem não ocorre ($\Delta=0$);
3. Supõe-se que os valores observados de $P_i(t)$ são confiáveis para idades menores que 30 ou 35 anos, mas, a partir daí podem estar afetados por erros de memória;
4. Em vez da sequência f_p, \dots, f_7 seria conveniente ter a informação para intervalos etários convencionais e, de preferência, no formato de uma função contínua f^t ;
5. A função de densidade $P^f(x, t)$ precisa ser determinada a partir dos dados pois geralmente estes especificam a população só em intervalos quinquenais.

Supondo que $f^+(y,t)$ seja uma função contínua de fecundidade derivada das observações f_p, \dots, f_7 , conforme a sugestão no ponto 4) acima, que corrige o problema da defasagem assinalado em 2), pode-se escrever a expressão seguinte, equivalente a (23.3)

$$F_i(t) = \int_{x_i}^{x_i+n} P^f(x,t) \int_{\alpha}^x f^+(y,t+y-x) dy dx / \int_{x_i}^{x_i+n} P^f(x,t) dx \quad (23.4)$$

Noutras palavras, F_i é o equivalente de P_i , mas baseado na fecundidade observada nos últimos 12 meses, tomando em conta o meio ano de defasagem mencionado no ponto 2). Como a fecundidade observada, inclusive a função $f^+(y,t)$, tem erros de nível, a estratégia geral consiste em usar P_i/F_i de um dos intervalos ($i=2$ ou $i=3$) onde o valor observado de P_i é relativamente recente e confiável como fator de correção para $f^+(y,t)$. Entretanto, para efetivamente calcular o fator P_i/F_i é preciso lidar com quatro questões:

1. Como fazer a conversão de f_p, \dots, f_7 em $f^+(y,t)$?
2. A função $f^+(y,t)$ em princípio só é conhecida no momento t , mas (23.4) exige que seja conhecida em $t+y-x$, para diferentes valores de x e y . Como essa generalização deve ser feita ?
3. Como lidar com o peso populacional $P^f(x,t)$ em (23.4) ?
4. Caso P_2/F_2 seja diferente de P_3/F_3 , como deve ser calculado o fator de correção ?

Existem várias abordagens para lidar com estas questões, a fim de obter uma função $f(y,t)$ corrigida, cada uma com os seus supostos e limitações específicos. Aqui se faz referência a quatro delas.

23.3.2.1 A abordagem original de Brass para a medição da fecundidade

Brass (1968, 1975) fez os seguintes supostos simplificadores:

1. A função $f(y,t)$ é constante no tempo, ou seja, $f(y,t) = f(y)$;
2. A função $f(y)$ é do tipo introduzido em (20.23), ou seja, $f(y) = C (y - s) (s + 33 - y)^2$;
3. Já que tanto o P_i como o F_i são ponderados pela distribuição da população $P^f(x,t)$ dentro do intervalo i , não faz muita diferença no cálculo de P_i/F_i supor que a distribuição é homogênea, ou seja, $P^f(x,t)$ é constante em x dentro de cada intervalo etário. Sendo assim, o fator $P^f(x,t)$ desaparece tanto de (23.3) como de (23.4).

Sob esses supostos, a função F_i iguala 5 vezes a soma de f_p, \dots, f_{i-1} mais uma parte de $5f_i$. A dificuldade reside em estimar qual é o multiplicador que deve ser aplicado a f_i . Para este fim, Brass calculou uma tabela, reproduzida abaixo como a Tabela 23.1, que permite obter os multiplicadores corretos a partir da relação observada entre f_1 e f_2 ou P_1 e P_2 ou a idade média (\bar{m}) da fecundidade recente para os demais grupos etários. Embora o padrão etário de (20.23) dependa apenas do

parâmetro s , (na verdade, não conhecido) Brass sugere, uma vez que o padrão real pode diferir do teórico, o uso de dois parâmetros: f_1/f_2 ou P_1/P_2 e \bar{m} para estimar os multiplicadores que devem ser aplicados a f_i e assim definir esse padrão real. Inicialmente, o autor utilizou f_1 e f_2 para os três primeiros grupos etários e \bar{m} para os seguintes; posteriormente ele recomendou substituir f_1/f_2 por P_1/P_2 , por considerar, este último, um parâmetro mais robusto.

Tabela 23.1 Fatores multiplicadores para se estimar a fecundidade corrente acumulada nos três primeiros anos dos grupos quinquenais com deslocamento de 0,5 ano (14,5 a 19,5, 19,5 a 24,5 etc.)

15-19 anos	1,120	1,310	1,615	1,950	2,305	2,640	2,925	3,170
20-24 anos	2,555	2,690	2,780	2,840	2,890	2,925	2,960	2,985
25-29 anos	2,925	2,960	2,985	3,010	3,035	3,055	3,075	3,095
30-34 anos	3,055	3,075	3,095	3,120	3,140	3,165	3,190	3,215
35-39 anos	3,165	3,190	3,215	3,245	3,285	3,325	3,375	3,435
40-44 anos	3,325	3,375	3,435	3,510	3,610	3,740	3,915	4,150
45-49 anos	3,640	3,895	4,150	4,395	4,630	4,840	4,985	5,000
f_1/f_2	0,036	0,113	0,213	0,330	0,460	0,605	0,764	0,939
\bar{m}	31,7	30,7	29,7	28,7	27,7	26,7	25,7	24,7
P_1/P_2^*	0,014	0,045	0,090	0,143	0,205	0,268	0,330	0,387

* O parâmetro P_1/P_2 foi cedido por Brass a alguns dos seus discípulos e ele passou a utilizá-lo por considerá-lo mais robusto (para mais detalhes, ver Carvalho, Gonçalves e Silva, 2018).

Fonte: Adaptado de Brass et al. (1968).

O passo final depois do cálculo do fator de correção P/F é a eliminação da defasagem em f_1, \dots, f_7 . Para este propósito se usam as seguintes fórmulas aproximadas:

$$f(15-19) = (f_1 + 0,05 f_2) P/F \quad (23.5.a)$$

$$f(20-24) = (f_2 - 0,05 f_1 + 0,05 f_3) P/F \quad (23.5.b)$$

.....

$$f(40-44) = (f_6 - 0,05 f_5 + 0,05 f_7) P/F \quad (23.5.f)$$

$$f(45-49) = (f_7 - 0,05 f_6) P/F \quad (23.5.g)$$

Embora a Tabela 23.1 seja a forma consagrada de aplicação do método original de Brass, é preciso observar que os números desta tabela também podem ser escritos explicitamente como fórmulas. As fórmulas são um pouco complicadas, mas hoje em dia não é particularmente difícil escrever um programa em “R” ou EXCEL para implementar o método pelas integrações relevantes da função (20.23). Isso tem as seguintes vantagens e acrescenta flexibilidade nos seguintes aspectos:

- Evita a necessidade de interpolações;
- Permite maior flexibilidade na escolha da função de fecundidade; por exemplo, a amplitude arbitrária de 33 anos na função (20.23) pode ser substituída por outro valor;

- A Tabela 23.1 parte de uma defasagem de 0,5 anos, por ser o caso mais comum. Outras defasagens possíveis, como 0 ou 1 ano, exigem outras tabelas. Um programa em EXCEL ou “R” pode contemplar estas diferentes possibilidades de uma forma mais flexível;
- Podem-se usar outros critérios de ajuste, além dos três critérios mostrados na Tabela 23.1. Por exemplo, em vez de usar apenas o critério $f_1/f_2 = 0,330$, que segundo a Tabela 23.1 leva a $k_1 = 1,950$, $k_2 = 2,840$ e $k_3 = 3,010$, seria possível exigir simultaneamente que $f_1/f_2 = 0,330$ e $f_2/f_3 = 0,95$, fazendo a amplitude de 33 anos varável, o que leva a $k_1 = 1,898$, $k_2 = 2,870$ e $k_3 = 3,070$.

Teoricamente também é possível relaxar o suposto 3), de homogeneidade da população dentro de cada intervalo etário. Entretanto, para que seja possível corrigir o viés de heterogeneidade (que de qualquer forma é pequeno), é preciso que a função de fecundidade esteja muito próxima a (20.23). Para funções de fecundidade de outros tipos, a correção da heterogeneidade da distribuição da população dentro de cada intervalo faz mais mal do que bem.

Embora o método original de Brass ainda continue sendo aplicado por alguns, há demógrafos, hoje em dia, que consideram a técnica superada, pelas seguintes razões:

1. A curva (20.23) é um modelo excessivamente simples para descrever o padrão etário da fecundidade a qual não se ajusta bem aos padrões observados, principalmente na África.
2. Em casos onde o padrão de fecundidade é significativamente diferente do modelo teórico de (20.23), o esquema de multiplicadores da Tabela 23.1 torna-se um pouco fictício, principalmente nas idades mais avançadas.
3. O suposto da constância da fecundidade no passado (ponto 1 acima) pode ter um efeito significativo sobre a série dos P_t , principalmente se afeta a fecundidade nas faixas de 15-19 e 20-24 anos. Isso muda as estimativas para países como o Brasil e Cabo Verde, onde a fecundidade caiu dramaticamente nas décadas passadas. Quando a queda no Brasil começou nos anos 70 e 80, Carvalho (1985) inicialmente argumentou que as mulheres afetadas geralmente eram mulheres mais velhas que já não queriam mais filhos, de modo que P_2 não mudava significativamente, mas as quedas mais recentes também afetam mulheres mais jovens (Castanheira e Kohler, 2015; Carvalho, Gonçalves e Silva, 2018). Neste caso, a fecundidade recente acumulada das mulheres mais jovens afeta o quociente P/F sobre-estimando o fator de ajuste a ser aplicado ao nível da fecundidade.

23.3.2.2O método do Manual X para a estimação da fecundidade

O método exposto no *Manual X* das Nações Unidas (1986) segue a mesma lógica geral do método original de Brass, com duas diferenças. Por um lado, em vez de adotar a função (20.23), usa os padrões de fecundidade de Coale e Trussell (20.28), que são mais flexíveis. Por outro lado, em vez da interpolação numa tabela, o método do *Manual X* usa uma abordagem de regressão com coeficientes padronizados, o qual é computacionalmente mais conveniente. Entretanto, o uso de

um modelo de fecundidade mais flexível também implica que é preciso empregar mais indicadores para determinar o padrão etário correto. A fórmula para o cálculo de F_i a partir de f_p , no caso de uma fecundidade recente defasada em 6 meses, é a seguinte:

$$F_i = \varphi_{i-1} + a(i)f_i + b(i)f_{i+1} + c(i)\varphi_7 \quad (23.6)$$

onde $\varphi_i = 5(f_1 + \dots + f_i)$ e os coeficientes $a(i)$, $b(i)$ e $c(i)$ são dados na Tabela 23.2. Nos extremos se usa $\varphi_0 = 0$ e, no último intervalo, f_6 em vez de f_7 .

Tabela 23.2: Multiplicadores propostos no *Manual X* para a estimação de F_i no caso onde a fecundidade recente está defasada em 6 meses¹

Idades	i	$a(i)$	$b(i)$	$c(i)$
15-19	1	2,531	0,188	0,0024
20-24	2	3,321	- 0,754	0,0161
25-29	3	3,265	- 0,627	0,0145
30-34	4	3,442	- 0,563	0,0029
35-39	5	3,518	- 0,763	0,0006
40-44	6	3,862	- 2,481	- 0,0001
45-49	7	3,828	0,016	- 0,0002

Fonte: Naciones Unidas (1986): Cap. II, Cuadro 7.

Tabela 23.3: Multiplicadores propostos no *Manual X* para a correção da defasagem de 6 meses

Idades	i	$x(i)$	$y(i)$	$z(i)$
15-19	1	0,031	2,287	0,114
20-24	2	0,068	0,999	- 0,233
25-29	3	0,094	1,219	- 0,977
30-34	4	0,120	1,139	- 1,531
35-39	5	0,162	1,739	- 3,592
40-44	6	0,270	3,454	- 21,497

Fonte: Naciones Unidas (1986): Cap. II, Cuadro 8.

O método das Nações Unidas também usa um procedimento um pouco mais preciso para eliminar a defasagem da série f_1, \dots, f_7 , caso haja necessidade. O análogo das fórmulas (23.5.a-g) é o seguinte:

$$f(15-19) = (f_1 + w(1)f_2) P/F \quad (23.7.a)$$

$$f(20-24) = ((1-w(1))f_2 + w(2)f_3) P/F \quad (23.7.b)$$

.....

$$f(40-44) = ((1-w(5))f_6 + w(6)f_7) P/F \quad (23.7.f)$$

$$f(45-49) = (1-w(6))f_7 P/F \quad (23.7.g)$$

¹ A publicação original também apresenta uma tabela para o caso onde a fecundidade recente provém de estatísticas vitais, ou seja, sem a defasagem de 6 meses.

onde

$$w(i) = x(i) + y(i) f_i / \varphi_7 + z(i) f_{i+1} / \varphi_7 \quad (23.8)$$

Os multiplicadores $x(i)$, $y(i)$ e $z(i)$ são mostrados na Tabela 23.3.

O método é aplicado automaticamente pelo módulo FERTPF do pacote MORTPAK e pelo PFRATIO de PASEX (ver seção 17.2 do Capítulo 17).

Os valores de P_i/F_i determinados pelo método de Brass ou das Nações Unidas permitem um diagnóstico do comportamento e a evolução da fecundidade. O que importa é tanto o nível de P/F , como a tendência. Um valor de 1 indica uma ausência de erros nas declarações de nascidos vivos. Caso a razão P/F seja menor que 1, pode ser um indicativo de:

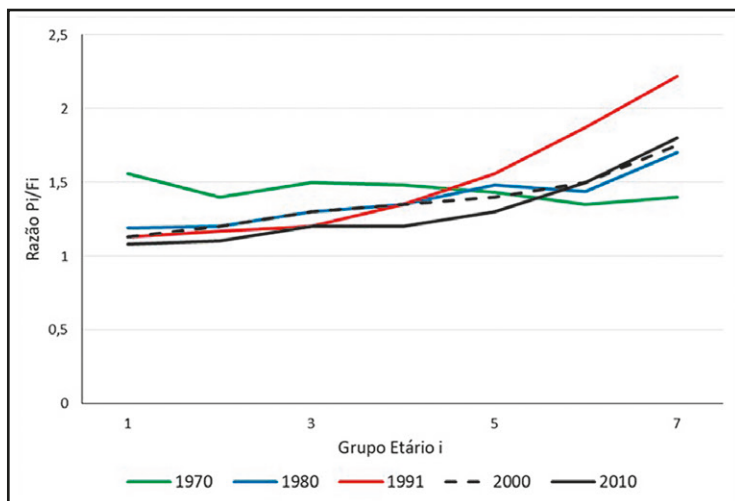
- a) Omissão do número de filhos nascidos vivos até o momento da pesquisa; e/ou
- b) Sobre-enumeração dos nascidos vivos no último ano (últimos 12 meses).

Por outro lado, caso a razão P/F seja maior que 1, as explicações podem ser:

- a) Omissão de nascidos vivos no último ano;
- b) Sobre-enumeração de nascidos vivos, talvez por inclusão dos que nasceram mortos;
- c) Uma fecundidade mais alta no passado.

Uma série P_i/F_i que aumenta com a idade i geralmente indica um declínio da fecundidade ao longo do tempo. Neste caso, é preferível usar P_2/F_2 para corrigir a fecundidade corrente, por ser uma informação mais recente. Se a série for mais constante, geralmente se usa a média de P_2/F_2 e P_3/F_3 .

Gráfico 23.2: Evolução da série P/F no Rio Grande do Norte urbano, 1970 a 2010



Fonte: Censos Demográficos do Brasil, 1970 a 2010.

O Gráfico 23.2 ilustra a evolução das séries P_i/F_i nos Censos do Rio Grande do Norte, de 1970 a 2010. Em 1970 a série era razoavelmente constante, num nível que variava entre 1,4 e 1,55. Isso sugere que a fecundidade anterior a 1970 já estava declinando e que havia uma subdeclaração dos nascimentos dos últimos 12 meses que podia ser corrigida com um fator de 1,45 (a média entre P_2/F_2 e P_3/F_3). Todas as outras séries, principalmente a de 1991, mostram aumento de P_i/F_i com a idade, como consequência da queda da fecundidade. Por outro lado, o grau de subdeclaração da fecundidade corrente parece ter diminuído, de modo que, usando P_2/F_2 , correções da ordem de 1,1-1,2 são indicadas para os Censos de 1980 a 2010. Neste particular exemplo, a sensível queda da estimativa do erro do período de referência, após 1970, deu-se em consequência da mudança do quesito referente à fecundidade corrente: em 1970, perguntou-se sobre o número de nascimentos vivos nos 12 meses anteriores à data de referência do Censo. Nos censos posteriores, inquiriu-se a data de nascimento do último filho nascido vivo.

Durante muitos anos, o método de *Manual X* foi o instrumento preferido para a estimação indireta da fecundidade. Entretanto, ele possui a mesma limitação do método original de Brass, de supor a constância da fecundidade no passado. Escolher P_2/F_2 como fator de correção atenua as distorções decorrentes de tendências de aumento ou diminuição, mas não as elimina inteiramente.

23.3.2.3 O modelo relacional de Gompertz

O modelo relacional de Gompertz, que se baseia no suposto de que a fecundidade segue o padrão de (20.34), implementa uma solução diferente para o problema dos erros de referência de tempo em f_p, \dots, f_7 . A parte mais importante da aplicação do método é o ajuste das séries P_i (baseada na fecundidade retrospectiva) e φ_i (baseada na fecundidade corrente)² pelo modelo teórico (23.9). A ideia é usar uma função de ajuste que serve tanto para P_i como para φ_i , embora φ_i esteja sujeito a erros de tempo de referência e P_i não. Os métodos descritos na seção 20.4 do Capítulo 20 não podem ser usados para este fim. Em vez disso busca-se um método que permita ajustar o padrão da fecundidade, sem depender do nível. A solução está no ajuste das razões da fecundidade acumulada até as idades de x e $x+5$ anos, para diferentes valores de x . Assim, partindo de (20.33) e (20.34), pode-se formular a seguinte relação em termos de $F(x)/F(x+5)$, uma quantidade que não depende da TFT:

$$\begin{aligned} z(x) &= -\ln\left(-\ln\left(\frac{F(x)}{F(x+5)}\right)\right) = -\ln\left\{-\ln\left[\exp\left(-\exp\left(-\alpha - \beta Y_p(x)\right)\right) + \exp\left(-\alpha - \beta Y_p(x+5)\right)\right]\right\} = \\ &= \alpha + \beta Y_p(x) - \ln(1 - \exp(\beta[Y_p(x) - Y_p(x+5)])) \end{aligned} \quad (23.9)$$

Os valores de $Y_p(x)$ encontram-se no Apêndice 6. No caso da série P_i , supõe-se que $Z_i = -\ln(-\ln(P_i/P_{i+l}))$ nos pontos $x_i = 17,5, 22,5, 27,5$ etc., ou seja, assume-se uma distribuição homogênea da população no intervalo. No caso de φ_i supõe-se que $Z_i = -\ln(-\ln(\varphi_i/\varphi_{i+l}))$ nos pontos $x_i = 19,5, 24,5, 29,5$ etc., tomando em conta o meio ano de defasagem. Se o erro de referência de tempo for o

² Optou-se pelo uso do símbolo φ_i , por razões de consistência com a simbologia usada anteriormente, em (23.5), mas no contexto do modelo relacional de Gompertz, esta, é comumente notada como F_i , o que cria certa confusão com o significado de F_i no contexto do método P/F tradicional.

mesmo para todos os grupos etários, as razões $F(x)/F(x+5)$ deveriam ser corretas, tanto para a série P_i , como para a série φ_i . O objetivo é encontrar os valores de α e β que minimizem

$$\sum_{\text{todos os pontos } P_i \text{ e } \varphi_i} (Z_i - \alpha - \beta Y_p(x_i) + \ln(1 - \exp(\beta[Y_p(x_i) - Y_p(x_i + 5)])))^2 \quad (23.10)$$

A expressão (23.10) parece configurar uma equação de regressão linear com parâmetros α e β , mas o problema é o terceiro termo na direita, que também depende de β . Zaba (1981) conseguiu linearizar a expressão, usando uma expansão da série de Taylor³ de (23.9), como função de β em volta de $\beta=1$, o que resulta na seguinte fórmula aproximada:

$$z(x) - e(x) \approx \alpha + \beta g(x) + \frac{1}{2} (\beta - 1)^2 c \quad (23.11)$$

onde as funções $g(x)$ e $e(x)$ são as seguintes:

$$g(x) = \frac{Y_p(x+5)\exp(Y_p(x)) - Y_p(x)\exp(Y_p(x+5))}{\exp(Y_p(x)) - \exp(Y_p(x+5))} \quad (23.12.a)$$

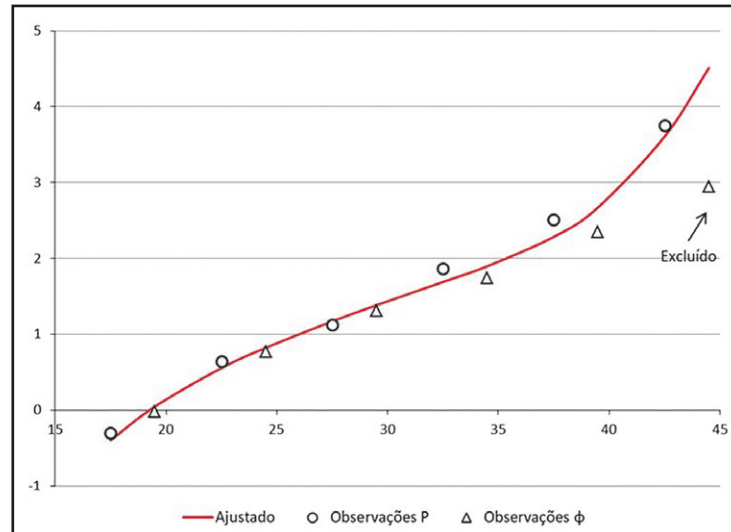
$$e(x) = -g(x) - \ln(\exp(-Y_p(x)) - \exp(-Y_p(x+5))) \quad (23.12.b)$$

A constante c em realidade não é constante, mas a segunda derivada de (23.9) em $\beta=1$, que empiricamente corresponde a uma função que quase não varia entre os 15 e 35 anos, com $c \approx 0,96$. Supondo que β é próxima de 1, o último termo de (23.11) pode ser omitido. Como as funções $g(x)$ e $e(x)$ dependem de $Y_p(x)$, mas não de α ou β , elas podem ser tabuladas, de modo que α e β podem ser determinadas com relativa facilidade, a partir de (23.11), usando uma regressão linear nos pontos 17,5, 19,5, 22,5 etc., que correspondem a P_p , φ_p , P_2 etc.

Embora (23.11) seja a forma consagrada de aplicação do método, usada inclusive por Moultrie et al. (2013), ela é, em realidade, desnecessariamente complicada. Na época em que Zaba publicou seu artigo, era quase impossível trabalhar diretamente com a fórmula (23.10), de modo que era preciso linearizá-la de alguma forma. Mas, hoje em dia, não existe essa limitação, de modo que acaba sendo mais fácil minimizar (23.10) diretamente, como função de α e β , por meio do comando *Solver* do EXCEL. Isso facilita tanto os cálculos como a intuição sobre o significado da relação. A única desvantagem desta forma de proceder é que o gráfico de (23.10) não é uma linha reta, o que faz a identificação visual dos desvios de Z_i em relação ao padrão esperado, um pouco mais difícil e não permite a estimação de α e β com as técnicas usuais de regressão linear.

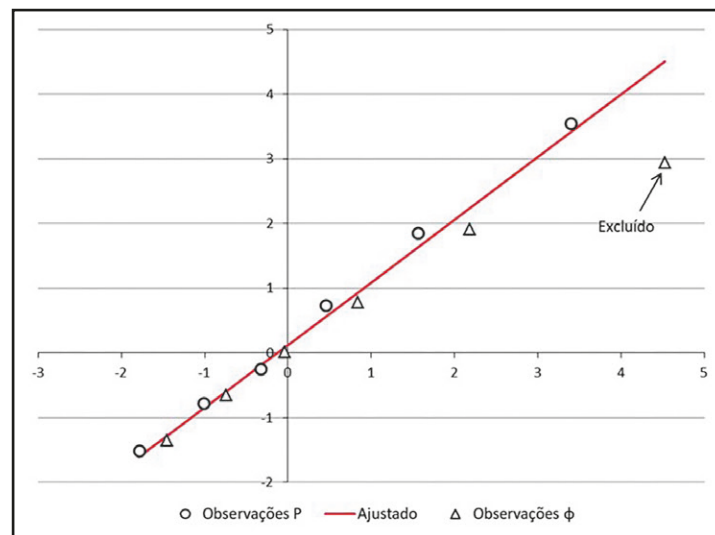
³ As séries de Taylor são um recurso poderoso no cálculo diferencial e integral, segundo o qual, sob certas condições de regularidade da função f e para um valor arbitrário de a , $f(x)$ pode ser aproximado por $f(x) = f(a) + f'(a)(x-a) + f''(a)(x-a)^2/2! + f'''(a)(x-a)^3/3! + \dots$. No caso da fórmula (23.10), $z(x)$, além de depender de x , também depende de α e β e foi representado como série de β , não de x , em volta do ponto $\beta=1$, até o terceiro termo.

Gráfico 23.3.A: Análise do padrão de fecundidade do Censo de Moçambique de 2017, usando a especificação exata



Fonte: Calculado com base nos resultados do Censo de Moçambique (2017).

Gráfico 23.3.B: Análise do padrão de fecundidade do Censo de Moçambique de 2017, usando a especificação linearizada



Fonte: Calculado com base nos resultados do Censo de Moçambique de 2017.

Ambas as variantes do método são ilustradas com o exemplo do Censo de Moçambique, de 2017, mostrado nos Gráficos 23.3.A e 23.3.B. O primeiro gráfico representa $z(x)$ como função da idade; a linha vermelha traça $z(x)$ como função contínua de x , para $\alpha = 0,1217$ e $\beta = 0,9678$. Os losangos azuis e triângulos verdes mostram os valores de Z_i para os pontos P_i e ϕ_i , respectivamente. O segundo gráfico mostra $z(x)-e(x)$ como função de $g(x)$. Os valores de α e β que caracterizam a linha reta em vermelho são quase idênticos aos valores do Gráfico 23.3.A: $\alpha = 0,1222$ e $\beta = 0,9677$. Estes valores podem ser obtidos pela aplicação das fórmulas usuais de regressão linear ou então da mesma forma como no Gráfico 23.3.A, pela minimização dos desvios quadrados em EXCEL, usando o Solver.

Em ambos os gráficos o ponto que corresponde a φ_7 foi excluído porque se distancia demais da tendência. Isso ilustra a recomendação geral de que os pontos nos quais se baseia a interpolação devem ser escolhidos com cuidado. Inicialmente todos os pontos são incluídos no modelo. A única exceção ocorre se a parturição média numa determinada faixa de idade for maior do que as parturições médias nos próximos grupos, por exemplo se $P(15-19)$ for maior do que $P(20-24)$. Neste caso, o gompito será indefinido e o modelo não pode ser ajustado usando esse ponto. Se há consistência entre os P_i e φ_i , todos os pontos se encontram perto da linha teórica. Caso contrário, os pontos que mais se afastam devem ser eliminados sequencialmente. A intenção é buscar a maior combinação de pontos P e φ que estejam (quase) na mesma linha, e usá-los para ajustar o modelo e encontrar α e β .

Moultrie (2013) dá algumas sugestões para auxiliar na seleção dos pontos:

- Uma série contínua de pontos deve ser incluída no modelo. Sequencialmente, apenas pontos nas extremidades podem ser excluídos.
- Pontos P devem ser eliminados de preferência. Isso ocorre porque parturições médias são geralmente mais propensas a erros específicos de idade do que os dados de fecundidade recente;
- Os pontos P e φ que divergem claramente da curva devem ser eliminados logo no início do processo de ajuste.
- Os pontos P e φ em idades mais velhas, como o ponto associado a φ_7 , devem ser eliminados de preferência, pois os dados nestas idades são geralmente os menos confiáveis e mostram menos coerência entre a vida reprodutiva e a fecundidade recente.
- Um ajuste marginalmente pior com mais pontos é preferível a um ajuste ligeiramente melhor com menos pontos.
- Um bom ajuste não deve sair das faixas de $-0,3 < \alpha < 0,3$ e $0,8 < \beta < 1,25$. Se os parâmetros estão fora destes intervalos, uma ou ambas as séries de dados subjacentes são problemáticas ou o padrão é inadequado. A alternativa é tentar outro padrão ou alterar a seleção de pontos, antes de prosseguir. Se os parâmetros ainda se situam fora dessas faixas, o método deve ser considerado inadequado.

O gráfico de pontos P e φ também serve como diagnóstico dos dados:

1. Caso ambas as sequências estejam numa linha reta, significa que os dados são perfeitos.
2. Omissão de filhos entre mulheres mais velhas. Nestes casos, um número errado de nascimentos (ou omissão) é reportado por mulheres mais velhas (P se eleva) e os valores de P tendem a ser maiores (denominador das φ acumuladas tende a ser menor).

3. Exagero na idade da maternidade, entre mulheres mais jovens. Mulheres jovens são classificadas como mais velhas do que são. Pontos ϕ caem nas idades mais velhas.
4. Tendências em fecundidade. Se a fecundidade cai, $\phi > P$ na mesma idade, e os pontos ϕ têm uma inclinação mais acentuada do que os pontos P . Os Pontos ϕ tendem a se alinhar acima dos pontos P . Assim, mudanças rápidas na fecundidade nas idades mais jovens impedem que os pontos ϕ e P fiquem alinhados, mesmo quando pontos P são excluídos.

Uma vez encontrados α e β e utilizando as fórmulas, $Y(x) = \alpha + \beta Y_p(x)$ e $Y(i) = \alpha + \beta Y_p(i)$, aplica-se uma transformação (inversa do gompito, $\exp(-\exp(-Y(x)))$) para encontrar a proporção de fecundidade corrente acumulada e proporção de fecundidade alcançada em cada grupo de idade. Dividindo a parturição média observada em cada faixa etária por essas proporções produzem uma série de estimativas de fecundidade total. Calculando a média destes valores em todo o subconjunto de faixas etárias utilizados para estimar α e β , obtém-se o valor ajustado estimado da fecundidade total.

A análise nos parágrafos anteriores ainda parte do suposto implícito de uma fecundidade constante ao longo do tempo. Zaba (1981) tentou flexibilizar este último pressuposto e estender o modelo Gompertz relacional para cenários nos quais a fecundidade muda no tempo. Para este propósito, faz-se uso de dois conjuntos de dados de parturição, obtidos em dois pontos no tempo. O modelo resultante é conhecido como o *modelo de Gompertz relacional sintético*.

Junto com informações de fecundidade corrente, fruto de informações de nascimentos recentes classificados por idade, o método permite flexibilizar o pressuposto de mudanças na fecundidade, podendo ser aplicado para censos ou pesquisas com 5 ou 10 anos de diferença. Com ele, também é possível identificar a coorte sobrevivente no primeiro período, por meio da informação dela num segundo momento. As mudanças na parturição média podem ser calculadas facilmente mediante o incremento de P_i entre as pesquisas. Uma sequência de incrementos de parturição (de diferentes coortes) entre as pesquisas pode ser estimada e isto resulta numa parturição média de *uma coorte hipotética*. Esta última, experimenta a fecundidade implícita pelos incrementos de parturição observados. As taxas de fecundidade de período são comparadas com as estimativas dessa coorte sintética. Estes devem idealmente se referir a todo o período entre os dois censos ou inquéritos que perguntam sobre a história reprodutiva.

Caso existam dados sobre nascimentos registrados, classificados pela idade da mãe, *para cada ano do período*, todos os nascimentos registrados, durante o período e em cada faixa etária, podem ser calculados pela adição por anos-calendário. As taxas médias de fecundidade para o período entre os inquéritos podem ser obtidas dividindo os nascimentos pelo número de mulheres-ano em cada faixa etária, estimadas a partir da população feminina enumerada no início e no final do período.

Caso os dados não estejam disponíveis, uma forma simples é estimar as TEFs por idade para o primeiro e último ano do período, e estimar as taxas para todo o período por meio de uma aproximação pela média aritmética entre os dois conjuntos de estimativas. Isso possibilita que o método seja aplicado em diferentes tipos de pesquisa, como inquéritos ou censos, com informações de nascimentos por idade da mulher.

Com base nas taxas médias e parturições de médias de uma coorte hipotética entre pesquisas, obtém-se uma sequência de pares de pontos φ e P , sujeitos aos mesmos procedimentos descritos no Gompertz relacional convencional, para obter taxas ajustadas e parturições médias intercensitárias.

Existem algumas condições para a aplicação do método:

1. A função de fecundidade padrão escolhida reflete a forma da distribuição de fecundidade da população observada, e quaisquer mudanças de fecundidade entre os censos ou inquéritos deve ser suave, gradual e proporcional em todos os grupos etários.
2. Os erros nas taxas, antes do ajuste, devem ser proporcionalmente os mesmos para as mulheres nas faixas etárias centrais (20-39).
3. O padrão etário da fecundidade, descrito por nascimentos relatados no passado, é razoavelmente preciso.
4. As parturições nas idades mais jovens (no início dos 20 anos de idade) são precisas.
5. As parturições médias da coorte sintética não diferem das parturições médias das mulheres que morreram ou migraram.

A aplicação do método também exige alguns cuidados. É importante que os dados sejam consistentes com relação à informação sobre classificação dos filhos pela idade da mulher. Por exemplo, no censo a informação referente a filhos nascidos nos últimos 12 meses tem uma defasagem de seis meses, enquanto o registro civil mostra a idade exata da mãe no momento do nascimento. Normalmente, não é bom combinar fontes, para evitar distorções.

Deve ser lembrado que apenas o *padrão* das taxas de fecundidade específicas por idade *entre pesquisas* é importante na aplicação do método de Gompertz relacional. Se esse padrão era mais ou menos constante ao longo do período, a data de referência das taxas usadas não importa. Mudanças no grau de cobertura por grupo de idade podem distorcer o padrão da fecundidade. Deve-se ter cuidado na aplicação nesses casos.

23.3.2.4 Um método que não exige um padrão explícito de fecundidade

Em parte o uso de modelos de fecundidade para a determinação das razões P/F se justifica pela necessidade de representar os cálculos necessários de uma forma simples e direta, como tabelas ou fórmulas de aplicação direta. Na época em que estes métodos foram formulados, nos anos 70 e 80 do século passado, esta necessidade se fazia sentir muito fortemente, tanto que às vezes se sobrepunha à exigência da precisão. Mas atualmente o problema da dificuldade computacional dos métodos é um impedimento muito menos importante, na medida em que se conta com instrumentos como EXCEL ou “R” para implementar os cálculos. Sendo assim, o uso de modelos de fecundidade como (20.23) ou (20.28), que reconhecidamente não descrevem bem o perfil de muitos países, particularmente os africanos, ou mesmo de (20.32) se tornou desnecessário.

Uma alternativa é o uso de splines para interpolar a função de fecundidade e o cálculo direto de F_i a partir das funções interpoladas. As vantagens deste procedimento são:

1. A função interpolada, em vez de refletir apenas algumas características da fecundidade observada, como f_1/f_2 ou P_1/P_2 e P_2/P_3 , pode reproduzir a sequência completa de f_1, \dots, f_7 .
2. A defasagem Δ de f_1, \dots, f_7 pode ser programada de forma flexível como sendo 0, 0,5, 1 ou qualquer outro valor, dependendo das necessidades.
3. Todos os indicadores necessários (a série F_i , a série P_i/F_i , a estimativa final de $f(x)$) podem ser calculados diretamente, sem a necessidade de calcular funções intermédias não intuitivas que dificultam a compreensão do procedimento.
4. É possível corrigir o viés introduzido pela heterogeneidade da distribuição etária das mulheres.
5. O método pode ser adaptado para situações onde a fecundidade não foi constante no passado.

A desvantagem do método é que exige o cálculo de uma interpolação com splines da função de fecundidade, o que pode ser trabalhoso. A planilha de EXCEL necessária para executar esse procedimento é demasiado extensa para ser reproduzida aqui, mas as suas características gerais são as seguintes:

1. Escolher um método de interpolação para a densidade de população por idade das mulheres. Existem diferentes maneiras de fazer esta interpolação. Aqui se escolheu um conjunto de 5 polinômios de terceiro grau para representar a população nas idades de (10,15), (15,20) e (20,25) anos (primeiro polinômio), (25,30) anos (segundo polinômio), (30,35) anos (terceiro polinômio), (35,40) anos (quarto polinômio) e (40,45) e (45,50) anos (polinômio final). Estes polinômios foram escolhidos para juntar-se de forma diferenciável nas idades de 25, 30, 35 e 40 anos.
2. Escolher as derivadas nas idades de 25, 30, 35 e 40 anos, de tal forma que as eventuais oscilações da interpolação sejam minimizadas.
3. Escolher oito polinômios de terceiro grau para representar $f(x)$ nos intervalos de (10- Δ , 15- Δ) a (45- Δ , 50- Δ) anos, onde Δ se refere à defasagem (0,5 anos, na maioria dos casos). As integrais destas funções, ponderados pela população de cada intervalo, devem corresponder às taxas de fecundidade observadas em cada intervalo e os polinômios devem se juntar de forma diferenciável nas idades de 15, ..., 45 anos. Para o primeiro polinômio se pode exigir que o seu valor seja igual a 0 numa idade inicial s , que pode ser 12, 13, 14 etc. anos, dependendo das circunstâncias. Para o último polinômio pode-se introduzir uma exigência de que assumam um valor de 0 na idade de 50, 51, 52, etc. anos, dependendo do caso.

4. Otimizar os parâmetros livres (derivadas em 15, ... , 45 e as idades inicial e final) para que o perfil da fecundidade seja o mais regular possível, com um mínimo de oscilações. Este passo costuma ser o mais difícil e, diferentemente de 1) e 3) exige a intervenção do pesquisador.
5. Deslocar a curva em meio ano para corrigir a defasagem Δ .
6. Calcular F_p , P_i/F_i e o perfil corrigido de fecundidade $f^*(x)$ a partir dos dados interpolados.

A Tabela 23.4 compara os resultados deste método (com e sem correção do viés de heterogeneidade da distribuição etário) com os métodos descritos nas seções 23.3.2.1 e 23.3.2.2, no caso do Censo de Moçambique de 2017.

Tabela 23.4: Estimação de F_i e $f(x)$ para o Censo de Moçambique (2017) com diferentes métodos, usando $P3/F3$ como fator de ajuste

	Splines com/sem correção do viés de heterogeneidade			
	Brass	Nações Unidas	Sem correção	Com correção
F_i				
10-14			0,0074	0,0074
15-19	0,2571	0,2379	0,2757	0,2763
20-24	1,0446	1,1332	1,0760	1,0689
25-29	1,9626	1,9748	1,9961	1,9774
30-34	2,7554	2,8096	2,7677	2,7650
35-39	3,4069	3,5011	3,4134	3,4049
40-44	3,7828	3,9153	3,8631	3,8511
45-49	4,1723	4,1173	4,1338	4,1240
f corrigido com base em P_i/F_i				
10-14			0,0126	0,0133
15-19	0,1460	,1552	0,1477	0,1503
20-24	0,2507	0,2513	0,2506	0,2534
25-29	0,2301	0,2281	0,2234	0,2241
30-34	0,1917	0,1891	0,1876	0,1893
35-39	0,1533	0,1515	0,1481	0,1484
40-44	0,0955	0,0887	0,0910	0,0908
45-49	0,0548	0,0511	0,0524	0,0523

*) Calculado usando o pacote MORTPAK, Versão 4 (ver seção 17.2 do Capítulo 17)

Fonte: Calculado com base nos resultados do Censo de Moçambique (2017).

Uma alternativa para o uso de splines é a interpolação com a distribuição normal modificada, descrita em (20.27) e $n=5$. Do ponto de vista da interpolação de f_p, \dots, f_p , este procedimento é, inclusive, mais simples do que o uso de splines. Entretanto, uma desvantagem do uso de (20.27) é que se trata de uma função que não pode ser analiticamente integrada para calcular a sequência F_p , de modo que é necessário recorrer a uma integração numérica.

23.3.2.5 Outros métodos para estimar a fecundidade quando ela não é constante

Como se viu acima, a aplicação dos métodos P/F , na sua forma convencional, exige a constância da fecundidade, pelo menos nos grupos etários mais jovens. Mas em países como o Brasil e Cabo Verde não é mais possível supor que este seja o caso, depois das quedas significativas havidas em décadas passadas, que afetaram todos os grupos etários. Uma das manifestações deste problema foi a constatação de que, no Brasil, os fatores P/F calculados para o Censo de 2010 foram maiores do que para 2000. Cavenaghi e Alves (2016) analisaram as possíveis causas e concluíram que se trata de uma mudança na estrutura da fecundidade, em decorrência de uma postergação da fecundidade das mulheres mais jovens, que afetou as razões P/F :

Se o ajuste da fecundidade for aplicado utilizando o valor de P/F do grupo de 20 a 24 anos de idade, o nível da TFT declarada no Censo, de 1,60 filho por mulher aumenta para 1,90, quando o nível da TFT registrada no Sinasc é de 1,72, empregando dados de 2009 e 2010 e a população feminina estimada para o meio do período. No entanto, como sabe-se que o Sinasc também não tem cobertura de 100% em todo território nacional, a aplicação do método P/F para ajustar esta fecundidade também seria recomendada. Para isso, além da população estimada de mulheres, é necessário usar os dados de parturição coletados no Censo Demográfico. Com este procedimento, a TFT estimada com dados de fecundidade corrente do Sinasc seria de 1,84 filho por mulher (...). Portanto, conclui-se que a estimativa da TFT a partir dos dados do Censo Demográfico de 2010 precisa do ajuste de cerca de 19% para um valor mais próximo do nível “verdadeiro”. Assim, a comparação destes dados indica que a fecundidade pode estar próxima de 1,9 e 1,84 filho por mulher, mas o verdadeiro valor não tem como ser conhecido, visto que o registro é deficiente e a fecundidade de período, calculada pelo Censo, pode estar afetada por mudanças conjunturais na estrutura da fecundidade e erros ou supostos que não se cumprem (Cavenaghi e Alves, 2016).

Diante desta situação, diferentes autores apresentam soluções distintas. Segundo Alves e Cavenaghi, ainda que seja válido buscar outras maneiras de estimar as taxas de fecundidade a partir de diferentes formas de ajuste que evitam os vieses do método P/F , todas as demais medidas apresentarão outros erros. Lima, Queiroz e Zeman (2018) experimentaram com três métodos: o método P/F de Brass, o método Gompertz Relacional Sintético aplicado a dois censos e o método de filhos próprios que será discutido abaixo, na seção 23.3.3. Eles concluíram que o método Gompertz Relacional Sintético apresenta os melhores resultados.

A solução mais radical foi sugerida por Castanheira e Kohler (2016) que consideram o método P/F inadequado, dada a dinâmica recente da fecundidade no país. Em vez de usar o fator P/F para ajustar o censo, eles propõem usar a informação do censo sobre a percentagem de crianças menores de 1 ano cujo nascimento foi registrado no SINASC (97,25%) como fator de correção para o SINASC. O resultado é uma TFT de 1,76, comparada com o intervalo de 1,84 a 1,90 encontrado por Cavenaghi e Alves e a estimativa oficial de 1,90 pelo IBGE.

A proposta de Carvalho, Gonçalves e Silva (2018) é bem mais conservadora e específica para o caso do Brasil e que se aplica a casos em que a fecundidade das mulheres de 15-19 anos diminuiu. Ela consiste utilizar, para corrigir os erros no período de referência dos dados de 2010, os fatores de ajuste P/F encontrados a partir da análise do Censo de 2000. Isso porque, para os autores, a estimativa do erro de período de referência, obtidos a partir das informações de 2000, é menos afetada pela variação de f_t do que aquela a partir de 2010, uma vez que a variação nos anos 1990 foi muito menor.

Moultrie e Dorrington (2008) fizeram simulações que sugerem que o método P/F pode sobre-estimar a fecundidade em até 5%, no caso de um forte declínio da fecundidade, além de apresentar

algumas outras limitações que causam diferentes tipos de alterações com a estimação por esse método em diferentes cenários. Para eles, a principal limitação seria mudanças nos níveis e na distribuição etária da fecundidade. Eles propõem um ajuste para o método, a partir da informação de dois censos (Nações Unidas, 1986) e, neste caso, o pressuposto de fecundidade retrospectiva constante é deixado de lado. Portanto, a fecundidade corrente pode ser comparada com a paridade acumulada de um ponto a outro.

Uma versão da metodologia para a estimação da fecundidade com base em dois censos foi originalmente desenvolvida por Arriaga (1983). Os passos do método, cuja lógica tem certa semelhança com a versão sintética do modelo relacional de Gompertz, são os seguintes:

1. Obter a parturição média por idade simples em cada censo ($P_{1,x}$ e $P_{2,x}$), interpolando os dados quinquenais por meio de um polinômio de 9º grau que reproduz os valores de P_t , que assume um valor de 0 em $x=15$ e cuja derivada é 0 em $x=15$ e $x=50$.
2. Agora o suposto básico do método é que a TEF entre a idade de x e $x+1$ anos no momento t_p pode ser interpolada linearmente entre t (data do primeiro censo) e $t+\Delta t$ (data do segundo censo), da seguinte forma:

$$1f_x(t_p) = ((\Delta t - t_p - 1/2)P_{1,x+1} + (t_p + 1/2)P_{2,x+1} - (\Delta t - t_p + 1/2)P_{1,x} - (t_p - 1/2)P_{2,x})/\Delta t \quad (23.13)$$

A ideia de que as TEFs nos anos entre os dois censos podem ser representadas por esta interpolação linear é um pressuposto muito forte do método. Implicitamente (23.13) supõe não só a linearidade da mudança da fecundidade, mas também assume que $P_{1,x+1} - P_{1,x}$ e $P_{2,x+1} - P_{2,x}$ representem a fecundidade na idade x nos anos t e $t+\Delta t$, o que estritamente só é correto no caso de uma fecundidade constante.

3. Calcular as TEFs por grupos etários quinquenais a partir das taxas anuais calculadas em 2) para $t_p = t + 1/2$ e $t_p = t + \Delta t - 1/2$.
4. Acumular as taxas quinquenais até as idades de 20, 25, 30 e 35 anos.
5. Acumular as TEFs quinquenais correspondentes à pergunta da fecundidade nos últimos 12 meses.
6. Calcular os fatores de ajuste P_i/F_i a partir de 4) e 5).
7. Escolher qual dos fatores de ajuste será usado e usar este fator para ajustar a fecundidade dos últimos 12 meses.

Leite (1982) aplicou este método ao caso do Brasil e estabeleceu que a TFT no Brasil tinha diminuído de 5,86 em 1941 a 5,67 em 1972 e 3,82 em 1980. Uma das vulnerabilidades do método é que erros na declaração do número de filhos nascidos vivos ou na idade das mulheres podem ter uma influência significativa sobre os resultados. Um problema mais fundamental é o passo 2) acima, com o suposto forte expresso em (23.13). O método pode ser executado automaticamente por meio do módulo ARFE-2 ou ARFE-3 do pacote PASEX.

Também existem propostas para adaptar o método P/F de um único censo, de tal forma que possa ser usada no contexto de mudanças da fecundidade ao longo do tempo. Em vez de assumir que $f(y, t+y-x) = f(y, t)$ em (23.2) e (23.3), como o faz o método convencional de Brass, Schmertmann et al. (2013) propuseram a seguinte relação:

$$f(y, t + y - x) = e^{r(x-y)} f(y, t) \quad (23.14)$$

Isso significa um suposto de que a fecundidade em todas as idades diminuiu uniformemente com um fator anual e^{-r} . No caso em que $r = 0$, isso se reduz ao suposto convencional de fecundidade constante. Supondo, como anteriormente, que $f(y, t)$ e $f^+(y, t)$ diferem apenas por um fator constante, é possível formular a seguinte relação:

$$5 (f_1 + f_2 + \dots + f_7) \frac{p(x, t)}{\int_{\alpha}^x f^+(y, t) dy} \approx TFT e^{r(x-\mu_x)} \quad (23.15)$$

onde

$$\mu_x = \frac{\int_{\alpha}^x y f^+(y, t) e^{r(x-y)} dy}{\int_{\alpha}^x f^+(y, t) dy} \quad (23.16)$$

ou seja, a idade média com que as mulheres que atualmente têm x anos tiveram os seus filhos. O importante de (23.15) é a constatação de que o lado esquerdo da fórmula pode não representar a TFT no momento do censo ou do inquérito, mas representa bastante bem qual foi a TFT μ_x anos antes. Esta ideia em realidade é bastante parecida com a proposta por Feeney (1996).

Em (23.15), é razoável supor que $P_1(t) \approx p(17,5, t)$, $P_2(t) \approx p(22,5, t)$, etc. A função $f^+(y, t)$ é construída usando os splines quadráticos descritos na seção 18.2.3 do Capítulo 18. Isso permite escrever a seguinte equação de regressão:

$$\ln \left(TFT \frac{P_i}{F_i} \right) = \beta_0 + \beta_1 (\mu_i - x_i) \quad (23.17)$$

onde $x_i = 22,5, 27,5, 32,5$ etc. e μ_i é a idade média em que mulheres do grupo etário i tiveram os seus filhos. Neste modelo a TFT acaba sendo igual a $\exp(\beta_0)$. O suposto mais forte do método é (23.14), ou seja, a igualdade da rapidez da queda da fecundidade em todos os grupos etários, de modo que a estrutura por idade da fecundidade se mantém constante. Este suposto é questionável, principalmente quando a fecundidade está diminuindo relativamente rápido.

O seguinte método não exige a homogeneidade do declínio da fecundidade em todas as faixas etárias, nem que esse declínio seja linear ou exponencial. Isso acrescenta flexibilidade, mas não chega a considerar todos os fatores potencialmente relevantes e é relativamente trabalhoso. Os passos do método são os seguintes:

1. Derivar uma função contínua para descrever a densidade da população feminina nas faixas etárias de 15-49 anos. Neste caso foi usado um polinômio de quarto grau.

2. Usando a densidade em 1), interpolar as TEFs dos últimos 12 meses f_p, \dots, f_7 para obter uma função contínua $f^*(x)$. Para este propósito se usou a curva normal modificada em (20.27), tomando em conta que a fecundidade observada é a média do ano passado e não corresponde exatamente ao ponto x na curva.
3. Estipular uma função f de fecundidade para o presente e o passado com a seguinte forma:

$$f(y, t + y - x) = f(y, t) \left(1 + r \left(\frac{x - y}{20}\right)^\delta\right) \left(1 + \lambda(y - 30) \left(\frac{x - y}{20}\right)^\delta\right) \quad (23.18)$$

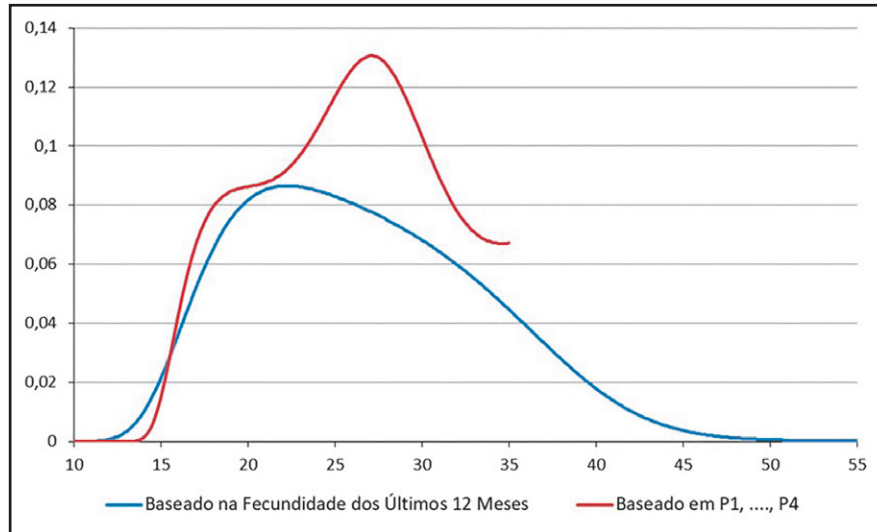
onde $f(y, t)$ é a fecundidade por idade y no momento do censo e $f(y, t + y - x)$ $x - y$ anos antes do censo. O parâmetro r indica a diferença entre o nível da fecundidade 20 anos antes e a atual. Ou seja, $r = 0,325$ significa que 20 anos antes do censo a fecundidade era 32,5% mais alta do que atualmente. O parâmetro δ permite modelar uma modificação não linear. Ou seja, $\delta = 1$ indica um perfil linear e $\delta = 2$ indica que o declínio inicialmente foi mais rápido e depois mais lento, como num polinômio de segundo grau. O parâmetro λ representa a mudança no padrão etário da fecundidade. O número 30 que aparece em (23.18) é o ponto de rotação do ajuste etário. Eventualmente se pode escolher um ponto diferente, como 27,5 anos. Se $\lambda > 0$, o padrão etário tem ficado mais jovem ao longo do tempo (maior nas idades abaixo de 30, menor acima dos 30); se $\lambda < 0$, o padrão etário tem ficado mais velho.

4. Aproximar $f(y, t)$ por meio de uma curva normal modificada do tipo (20.27) que depende da sequência observada P_1, P_2, P_3 e P_4 e dos parâmetros especificados em 3) por meio da relação (23.3). O resultado também depende da densidade da população subjacente calculada em 1. É de notar que o procedimento é o oposto daquele adotado no método convencional de Brass ou das Nações Unidas, mas é parecido com o procedimento usado no método de Arriaga: em vez de construir um indicador a partir de f_p, \dots, f_7 que possa ser comparado com P_p , se converte a série P_1, P_2, P_3, P_4 numa função de fecundidade que possa ser comparada com f^* . Se for usado um polinômio de quinto grau ($n=5$) para $f(y, t)$, não há informação suficiente para estimar todos os coeficientes. Por isso se acrescenta a condição de que, a partir dos 35 anos, o *padrão etário* de $f(x, t)$ deve ser igual ao padrão etário de $f^*(x)$.

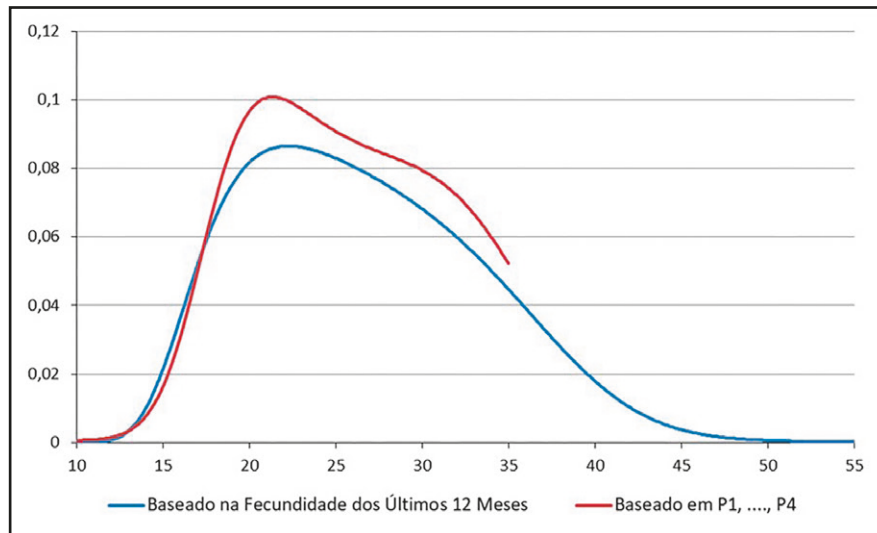
Teoricamente seria possível estimar os parâmetros r , δ e λ a partir da condição de que $f(x, t)$ deve ser proporcional a $f^*(x)$, mas na prática este procedimento se torna numericamente instável. Por isso é melhor partir de uma estimativa razoável de r , δ e λ – ou pelo menos de dois destes três parâmetros – e depois verificar se surge um padrão etário de $f(x, t)$ compatível com $f^*(x)$, ajustando o terceiro parâmetro, no caso de ser livre. Os procedimentos descritos acima são relativamente complexos – em parte porque (23.18) não pode ser analiticamente integrado e requer uma integração numérica – e precisam ser implementados num programa especial em algum aplicativo de cálculo com facilidades de otimização, como EXCEL ou “R”, para serem executados.

O Gráfico 23.4.A mostra os resultados para o caso onde $r = 0$, ou seja, o suposto subjacente à aplicação convencional do método P/F. O forte aumento da linha baseada em P_i por volta dos 30 anos indica que a fecundidade no passado foi substancialmente maior do que a atual. Usando apenas $(P_1 + P_2) / (F_1 + F_2)$ para corrigir a fecundidade corrente, obtém-se uma TFT de 1,89, muito

próxima da estimativa oficial de 1,90. No Gráfico 23.4.B os resultados são corrigidos como o suposto de que $r = 0,35$, $\lambda = -0,0325$ e $\delta = 2$. Com esses parâmetros o padrão etário das duas curvas coincide muito mais. A TFT corrigida acaba sendo 1,814. Grosso modo, isso confirma os resultados de Schmertmann et al.

Gráfico 23.4.A: Aplicação do método P/F modificado com $r = 0$ 

Fonte: Elaborado a partir do Censo Demográfico do Brasil de 2010.

Gráfico 23.4.B: Aplicação do método P/F modificado com $r = 0,35$, $\lambda = -0,0325$ e $\delta = 2$ 

Fonte: Elaborado a partir do Censo Demográfico do Brasil de 2010.

Vale mencionar ainda que o método pode ser adaptado para o caso de dois censos, com ou sem o uso da informação sobre filhos nascidos vivos durante os 12 meses antes do primeiro censo. Em ambos os casos, é preciso especificar o parâmetro δ que determina o perfil da modificação da fecundidade no tempo. Se a informação sobre filhos nascidos vivos no primeiro censo não é usada,

o método produz uma função de fecundidade corrente para o primeiro censo e um fator de ajuste P/F para o segundo. Neste caso, o método é parecido com o de Arriaga, mas sem o pressuposto forte (23.13). Caso contrário, produz fatores de ajuste P/F para ambos os censos.

23.3.3 O Método de Filhos Próprios

O método de filhos próprios (Cho e Feeney, 1978; Cho, Retherford e Choe, 1986; Wong, 1983; Fernández e Carvalho, 1986; Miranda-Ribeiro, Rios-Neto e Carvalho, 2009) permite fazer estimativas de fecundidade para um período que se estende no passado por mais ou menos 15 anos. A ideia por detrás do método é simples, embora a sua execução possa ser bastante trabalhosa. Se $P_{x,y}$ indica o número de crianças de x anos completos de idade cujas mães têm y anos, com uma soma de P_y para filhos de todas as idades, a Taxa Específica de Fecundidade no período entre x e $x+1$ anos antes do momento do censo, para mulheres que naquele momento tinham entre $y-x-1$ e $y-x$ anos deve ter sido

$$TEF_{y-x}(t - x - 1/2) = P_{x,y} \frac{\ell_0}{\ell_{x+1/2}} / P_y \frac{\ell_{y-x}}{\ell_{y+1/2}} \quad (23.19)$$

A lógica desta expressão é que os $P_{x,y}$ filhos atualmente vivos são os sobreviventes daqueles que nasceram na época. Para saber quantos foram naquele momento, é preciso multiplicar por $\ell_0 / \ell_{x+1/2}$. Da mesma forma, as P_y mães atuais são as sobreviventes das $P_y \ell_{y-x} / \ell_{y+1/2}$ que havia na época. Dividindo os dois números obtém-se a TEF do período para a idade correspondente das mulheres. Fazendo o mesmo para outras idades de mulheres e filhos é possível reconstruir todo um perfil de fecundidade para os últimos 10 ou 15 anos. O mesmo procedimento também pode ser aplicado separadamente para cada sexo.

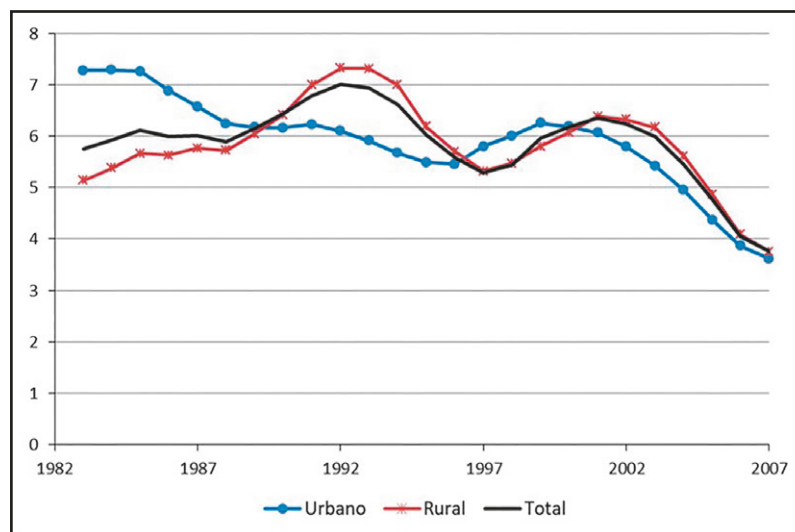
As dificuldades práticas na execução do método são as seguintes:

1. A fórmula (23.19) exige que se conheça a tábua de vida da população. O fato de que a razão de sobrevivência aparece tanto no numerador como no denominador faz o resultado menos sensível ao nível de mortalidade; se bem é verdade que erros na estimativa da sobrevivência de mulheres em idade reprodutiva afetariam minimamente o denominador, isso pode não ser o caso entre a população infantojuvenil, principalmente se a mortalidade infantil for alta ou estiver em acentuado declínio. Portanto, é preciso dispor de uma estimativa bastante razoável da mortalidade antes de aplicar o método.
2. Na prática pode não ser particularmente fácil juntar os filhos com as suas mães para saber qual é a sua idade. Alguns censos, incluindo o Censo brasileiro de 2010, e com frequência as pesquisas tipo DHS, contêm uma pergunta que pede a identificação da linha do questionário para a mãe de cada filho, ou seja, “a mãe deste filho é aquela que aparece na linha x do questionário” o que ajuda muito na execução do processo. Há outros critérios que podem ser usados, como a relação da criança com o(a) chefe do domicílio (agregado familiar), a idade de cada mulher, quantos filhos ela teve e se ela teve um filho nos últimos 12 meses, mas certamente existem situações – quando há várias mulheres adultas no mesmo

domicílio (agregado familiar) – onde não se pode afirmar com certeza quem é a mãe de cada filho. De todos os modos, a junção de cada filho à sua mãe normalmente exige um programa especial de processamento que precisa ser aplicado aos microdados do censo⁴.

3. Aqueles filhos que não vivem com as suas mães não podem ser juntados a elas. Normalmente se supõe que as idades das suas mães se distribuem da mesma forma que as idades das mães que sim puderam ser juntadas aos seus filhos.
4. Como o método descansa fundamentalmente na informação sobre população recenseada, as estimativas de fecundidade serão profundamente afetadas pelo grau de cobertura do censo utilizado. Dado o típico padrão de erro de cobertura censitária, que costuma ser muito mais alto entre as crianças menores de 2 ou 3 anos de idade e, muito mais acentuado ainda, entre os menores de um ano, é também, muito frequente, obter níveis de fecundidade subestimados provenientes da informação correspondente aos últimos anos antes da data do levantamento censitário. Isso pode sugerir um forte declínio da fecundidade recente que não corresponde à realidade.
5. Igualmente, a possibilidade de filhos de 10-15 anos não residir no mesmo domicílio da mãe dificulta a estimativa da idade a mãe, do cálculo da fecundidade, o que faz que estimativas de fecundidade muito distantes da época de realização do censo (10 ou 15 anos atrás) sejam, frequentemente subestimadas.
6. Como já foi mencionado no Capítulo 16, a qualidade das estimativas pode ser afetada por erros na declaração das idades, tanto dos filhos como das mães.

Gráfico 23.5: Fecundidade (TFT) total, urbana e rural em Moçambique 1983-2007, calculada com base no método de filhos próprios a partir dos Censos de 1987, 1997 e 2007



Fonte: Arnaldo (2014): Figure 16.4.

⁴ Usuários das bases de dados do IPUMS (ver Capítulo 4) podem aproveitar um recurso desenvolvido pela Universidade de Minnesota que tentativamente identifica o pai, a mãe e o/a cônjuge de cada membro do domicílio (se reside no mesmo domicílio ou agregado familiar) mediante “variáveis pontadores” (Sobek e Kennedy, 2009).

O Gráfico 23.5 ilustra bem os vieses típicos achados ao aplicar o método dos filhos próprios. As oscilações das TFT, antes que indicar a instabilidade de fecundidade no período de mais de 20 anos (1983-2007) são consequência, em grande parte, da subcobertura dos Censos de 2007, 1997 e 1987.

Como se mencionou acima, no ponto 2, um dos aspectos mais difíceis da aplicação do método, na sua forma “clássica”, é a necessidade de uma junção individual de cada mãe com os seus filhos e a tabulação cruzada dos filhos pela sua idade e a idade da sua mãe. Isso exige um processamento relativamente complexo da informação censitária. Entretanto, o processo pode ser simplificado significativamente ao aceitar alguma perda de informação. Timaeus e Moultrie (2013) observam que a TBN e a TFG podem ser calculadas sem a necessidade de uma junção individual dos dados de mães e filhos. Tudo que se precisa para este fim é o número de filhos de x anos exatos, inversamente projetados para obter o tamanho inicial da coorte, e o tamanho da população total (no caso da TBN) ou da população de mulheres em idade reprodutiva (no caso da TFG) x anos atrás. Basta ter a distribuição etária de mães e filhos separadamente, no momento do censo, e uma tábua de vida aproximada para retroprojetar estes números, sem a necessidade de tabulações especiais do censo. Se, além disso, existe alguma informação sobre a evolução do *padrão etário* (não o nível) da fecundidade durante os últimos 15 anos, por exemplo a partir das perguntas de nascimentos durante os últimos 12 meses em dois ou mais censos sucessivos, a sequência de TFGs pode ser convertida numa sequência de TEFs e, portanto, TFTs. Alguma informação será perdida no processo porque a evolução do padrão etário pode não ter sido tão suave como o método supõe, mas se o único objetivo for uma estimativa da evolução das TFTs, esta perda certamente é aceitável. Para maiores detalhes sobre a aplicação desta variante do método, consulte-se Timaeus e Moultrie (2013).

23.4 MÉTODOS PARA ESTIMAR A MORTALIDADE INFANTOJUVENIL

Os métodos mais comuns para estimar a mortalidade infantojuvenil se baseiam na ideia mais geral de medição da mortalidade via relações de parentesco. Embora essa classe de métodos seja aplicada mais frequentemente à mortalidade infantojuvenil, ela pode ser aplicada também à mortalidade adulta, principalmente pela análise da sobrevivência das mães dos respondentes.

Estimativas baseadas nas relações de parentesco foram propostas por Brass (1968, 1975). A ideia básica por trás é usar a informação dos filhos sobreviventes segundo a idade (ou o tempo desde a primeira união) da mãe e relacioná-la com a probabilidade de morte desde o nascimento até a idade x (${}_xq_0$). Para entender a lógica desse método, suponha que todos os nascimentos numa população ocorrem na idade materna exata de 19,5 anos. Então, se for perguntado às mulheres desta população sobre a sobrevivência dos seus filhos, a proporção média de filhos mortos por mulher numa idade exata x seria igual à probabilidade de morte de crianças desde o nascimento até a idade exata $x-19,5$. Noutras palavras, a proporção média de filhos sobreviventes de mulheres da idade exata x seria igual a $\ell_{x-19,5}$.

Aqui o tema será tratado com base nas idades atuais das mães (x), mas em populações onde a idade da primeira união é bem definida e onde essa idade efetivamente marca o início da reprodução (ou seja, poucos nascimentos antes da primeira união e pouco adiamento da reprodução dentro do matrimônio) é possível obter resultados mais confiáveis classificando as mães pelo tempo desde a primeira união. O módulo CEBCS de MORTPAK usa essa metodologia, enquanto QFIVE se baseia nas idades das mães, da mesma forma como será feito aqui.

Se todas as mulheres tivessem os seus filhos numa mesma idade bem definida y , todos os filhos de mulheres de idade x ($x > y$) teriam o mesmo tempo de exposição $x-y$ ao risco de morte. Numa população real, a relação se torna mais difusa na medida em que os filhos de mulheres de idade x não nasceram todos no mesmo momento, mas espelhados pelo tempo segundo a função de fecundidade $f(y,t)$. Esta função não foi necessariamente constante ao longo do tempo. Se o seu valor para mulheres que atualmente têm x anos for $f(x,t)$, então na época quando a mulher com x anos na atualidade tinha y anos, o seu valor era igual a $f(y,t+y-x)$. Isso leva à seguinte expressão para descrever a proporção média $s_x(t)$ de filhos sobreviventes de mulheres da idade exata x :

$$s_x(t) = \int_{\alpha}^x f(y, t + y - x) \ell_{x-y}(t + y - x) dy / \int_{\alpha}^x f(y, t + y - x) dy \quad (23.20)$$

onde α representa a idade inicial da fecundidade. Os supostos implícitos a (23.20) são que

1. As mulheres que morreram (e que portanto não foram entrevistadas) têm as mesmas funções f_y que as mulheres que sobreviveram e os seus filhos têm a mesma função ℓ_{x-y} . Tradicionalmente esse suposto parecia razoável, mas nas estimativas da mortalidade africana depois da década de 80 isso se tornou mais problemático devido à influência da AIDS (SIDA) que afeta a sobrevivência de tanto as mulheres como os seus filhos. O tema não será abordado sistematicamente aqui, mas é discutido, por exemplo, em Quattrochi et al. (2019);
2. As idades das mulheres são corretamente declaradas;
3. Não há erros de seletividade nas respostas das mulheres;
4. A população é fechada à migração; e
5. A mortalidade dos filhos não depende da idade das suas mães.

Os problemas mais comuns que surgem na aplicação do método se relacionam com 3) e 5). Como já foi assinalado anteriormente, as mulheres de mais de 35 anos às vezes não mencionam os filhos que nasceram há muito tempo e que morreram quando muito jovens. Isso causa um viés para baixo no número de nascimentos, mas principalmente no número de filhos mortos. O outro problema que muitas vezes ocorre diz respeito a mães muito jovens (15-19 anos) cujos filhos podem ter uma mortalidade acima do normal, seja como consequência direta da sua maternidade precoce ou porque as mães mais jovens tendem a ser desproporcionalmente mais pobres e ter níveis educacionais mais baixos (ver Capítulo 10).

Da mesma forma como acontece na análise da fecundidade, $s_x(t)$ não é observada para idades exatas x , mas para intervalos etários quinquenais. Portanto, o equivalente de (23.3) é a seguinte expressão, que quantifica a proporção $D_i(t)$ de filhos mortos de mulheres de 15-19 anos ($i=1$), 20-24 anos ($i=2$) etc. no momento t do censo ou do inquérito.

$$D_i(t) = 1 - \int_{x_i}^{x_{i+1}} Pf(x, t) \int_{\alpha}^x f(y, t + y - x) \ell_{x-y}(t + y - x) dy / \int_{x_i}^{x_{i+1}} Pf(x, t) \int_{\alpha}^x f(y, t + y - x) dy dx \quad (23.21)$$

O denominador da expressão é conhecido porque representa o número médio de filhos nascidos vivos (independentemente da sua sobrevivência depois do nascimento) de uma mulher de exatamente x anos, que é um dos dados recolhidos no censo ou no inquérito. O numerador contém a mesma expressão, só que ponderada pelo fator $\ell_{x-y}(t+y-x)$. À primeira vista, a fórmula parece demasiado complexa para resolver com facilidade e descobrir a forma da tábua de vida $\ell_x(t)$. Com os recursos computacionais disponíveis hoje em dia, não é impossível usar (23.3) e (23.4) para estimar $f(y,-t)$ e a partir daí estimar $\ell_x(t)$ em (23.21), mas na época em que a maioria dos métodos indiretos foi desenvolvida isso não era factível e mesmo na atualidade raramente é feito. Em vez disso usam-se várias soluções aproximadas para (23.21), da mesma forma como no caso de (23.3) e (23.4), para simplificar o procedimento, supondo que algumas condições simplificadoras estejam satisfeitas.

23.4.1 A abordagem original de Brass para a medição da mortalidade

Todas as variantes do método de estimação da mortalidade infantojuvenil se baseiam em (23.20); as diferenças dizem respeito aos supostos simplificadores adicionais usados para possibilitar a estimação de ℓ_{x-y} . No caso da variante proposta por Brass (1968, 1975) esses supostos adicionais se referem aos seguintes pontos:

1. Como todos os outros autores, Brass supõe que f é constante no tempo.
2. Brass assume que f tem a forma descrita em (20.23), que depende apenas do parâmetro s .
3. Brass assume um padrão de mortalidade baseado nas tábuas de vida modelo das Nações Unidas (1955), com algumas modificações; outros métodos se baseiam nas tábuas de vida regionais de Princeton ou das Nações Unidas (ver seção 20.2.3).
4. O método original proposto por Brass também supõe a constância da mortalidade dos filhos ao longo do tempo; alguns desenvolvimentos posteriores relaxam esse suposto.
5. Todos os métodos também supõem ou seja, $P^f(x,t)$ é constante em x dentro de cada intervalo etário, o que faz com que esse fator desapareça no numerador e denominador de (23.21).

Sob esses supostos, sabe-se matematicamente que cada D_i corresponde à probabilidade de morte entre o nascimento e uma determinada idade m_i , que é próxima à idade média dos filhos cujas mães encontram-se no grupo etário i , ou seja $D_i \approx_{m_i} q_0$. Brass verificou que para $i=1$, esta idade média é aproximadamente 1 ano, ou seja, $D_1 \approx_1 q_0$. Da mesma forma, $D_2 \approx_2 q_0$, $D_3 \approx_3 q_0$, $D_4 \approx_4 q_0$, $D_5 \approx_5 q_0$, $D_6 \approx_6 q_0$ e $D_7 \approx_7 q_0$. Embora estas relações sejam aproximadamente verdadeiras, a correspondência não é perfeita. Por exemplo, num país com fecundidade muito precoce os filhos de mulheres de um determinado grupo etário serão, na média, um pouco mais velhos do que num país de fecundidade mais tardia. O padrão de mortalidade também pode influenciar esta relação, embora em menor medida. Todos os métodos indiretos de estimação da mortalidade infantojuvenil partem da relação $D_i \approx_{m_i} q_0$, mas fazem ajustes para garantir que ela seja exata, em vez de apenas aproximada.

No caso da variante proposta por Brass, o autor introduz fatores de correção k_i , de tal forma que

$$k_i = q(m_i) / \{1 - \int_{x_i}^{x_{i+1}} \int_{\infty}^x f(y, 0) \ell_{x-y}(0) dy / \int_{x_i}^{x_{i+1}} \int_{\infty}^x f(y, 0) dy dx\} \quad (23.22)$$

Como no caso da fecundidade, Brass (1968) calculou uma tabela que permite obter o multiplicador k_i correto a partir da relação observada entre P_1 e P_2 ou P_2 e P_3 ou da idade média da fecundidade recente \bar{m} . Interpolando estes fatores de correção, se obtém a probabilidade correta como ${}_m q_0 = k_i D_i$.

Sullivan (1972) simplificou o cálculo de k_i ao fornecer equações de regressão explícitas para k_2 , k_3 e k_4 que variam segundo o modelo de mortalidade de Princeton que se assume:

$$\begin{aligned} \text{Oeste:} \quad k_2 &= 1,30 - 0,54 P_2 / P_3 \\ k_3 &= 1,17 - 0,40 P_2 / P_3 \\ k_4 &= 1,13 - 0,33 P_2 / P_3 \end{aligned} \quad (23.23.a)$$

$$\begin{aligned} \text{Norte:} \quad k_2 &= 1,30 - 0,63 P_2 / P_3 \\ k_3 &= 1,17 - 0,50 P_2 / P_3 \\ k_4 &= 1,11 - 0,42 P_2 / P_3 \end{aligned} \quad (23.23.b)$$

$$\begin{aligned} \text{Leste:} \quad k_2 &= 1,26 - 0,44 P_2 / P_3 \\ k_3 &= 1,14 - 0,33 P_2 / P_3 \\ k_4 &= 1,15 - 0,26 P_2 / P_3 \end{aligned} \quad (23.23.c)$$

$$\begin{aligned} \text{Sul:} \quad k_2 &= 1,33 - 0,61 P_2 / P_3 \\ k_3 &= 1,20 - 0,44 P_2 / P_3 \\ k_4 &= 1,14 - 0,32 P_2 / P_3 \end{aligned} \quad (23.23.d)$$

O suposto de constância da fecundidade não pode ser relaxado sem incorrer em inconsistências. Por exemplo, se a fecundidade no passado foi mais elevada, isso significa que os filhos de mulheres no grupo etário i nasceram, na média, há mais tempo do que sugerido pelo suposto de fecundidade constante, de modo que a mortalidade precisa ser ajustada para baixo. Mas (23.23.a-d) sugerem o oposto, pois um fecundidade em declínio deprime P_2/P_3 e portanto, aumenta os fatores de ajuste k_i . Apesar da existência deste problema, as distorções resultantes da violação do suposto de constância da fecundidade geralmente não são grandes. Por exemplo, supondo um padrão de mortalidade Oeste com $e_0 = 66,5$ anos e uma fecundidade de acordo com (20.23) com $s=14$ e um ritmo rápido de redução da fecundidade de 3% ao ano em todas as idades, (23.23.a) sobre-estima a mortalidade em apenas 2,8%.

Por outro lado, a mudança no nível de mortalidade pode ser acomodada de forma mais natural pelo fato de que os métodos de Brass e Sullivan geram estimativas de ${}_m q_0$ para diferentes idades, que podem ser interpretadas como níveis que se aplicam a diferentes momentos do passado. Feeney (1980) procurou alocar no tempo as estimativas de probabilidade de morte e verificou

que as funções de mortalidade, independente da tendência, passam por um mesmo ponto de ${}_2q_0$ em determinado momento do tempo. Isso indica que a cada medida de mortalidade de coorte existe uma medida equivalente de período t anos anteriores à data da realização da pesquisa. Portanto, combinando as estimativas de ${}_1q_0, {}_2q_0, {}_3q_0, {}_5q_0$ etc. com um sistema de tábuas de vida modelo, obtém-se uma sequência de níveis de mortalidade em diferentes momentos do passado.

23.4.2 O método do Manual X para a estimação da mortalidade

O método proposto no *Manual X* das Nações Unidas fez algumas adaptações ao método de Brass. Em vez de usar o modelo de fecundidade de Brass, usou os modelos de Coale-Trussell descritos em (20.28). Em vez do modelo de mortalidade das Nações Unidas (1955), usou as tábuas de vida modelo de Princeton (1983). Finalmente, as estimativas das Nações Unidas já incorporam as defasagens de tempos t_i calculados por Feeney nas próprias fórmulas. Como os modelos de Coale-Trussell têm mais graus de liberdade do que o modelo de fecundidade de Brass, as fórmulas do *Manual X* têm dois termos para especificar os fatores de correção, em vez de um. Abaixo se reproduzem apenas os coeficientes para o caso do modelo Oeste de Princeton. Os coeficientes para os demais modelos de Princeton podem ser encontrados nas Tabelas 47 e 48 do *Manual X*.

$$\begin{aligned}
 k_1 &= 1,1415 - 2,7070 P_1/P_2 + 0,7763 P_2/P_3 & t_1 &= 1,0970 + 5,5628 P_1/P_2 - 1,9956 P_2/P_3 \\
 k_2 &= 1,2563 - 0,5381 P_1/P_2 - 0,2637 P_2/P_3 & t_2 &= 1,3062 + 5,5677 P_1/P_2 + 0,2962 P_2/P_3 \\
 k_3 &= 1,1851 + 0,0633 P_1/P_2 - 0,4177 P_2/P_3 & t_3 &= 1,5305 + 2,5528 P_1/P_2 + 4,8962 P_2/P_3 \\
 k_4 &= 1,1720 + 0,2341 P_1/P_2 - 0,4272 P_2/P_3 & t_4 &= 1,9991 - 2,4261 P_1/P_2 + 10,4282 P_2/P_3 \\
 k_5 &= 1,1865 + 0,3080 P_1/P_2 - 0,4452 P_2/P_3 & t_5 &= 2,7632 - 8,4065 P_1/P_2 + 16,1787 P_2/P_3 \\
 k_6 &= 1,1746 + 0,3314 P_1/P_2 - 0,4537 P_2/P_3 & t_6 &= 4,3468 - 13,2436 P_1/P_2 + 20,1990 P_2/P_3 \\
 k_7 &= 1,1639 + 0,3190 P_1/P_2 - 0,4435 P_2/P_3 & t_7 &= 7,5242 - 14,2013 P_1/P_2 + 20,0162 P_2/P_3
 \end{aligned}$$

(23.24)

Enquanto o *Manual X* se limita a apresentar os coeficientes aplicáveis aos quatro modelos de Princeton, Palloni e Heligman (1986) estimaram os coeficientes para os cinco modelos de mortalidade desenvolvidos pelas Nações Unidas (Naciones Unidas, 1983). Os resultados não serão apresentados aqui. Eles podem ser encontrados no artigo de Palloni e Heligman ou em United Nations (1990). O programa QFIVE do pacote MORTPAK aplica tanto o método do *Manual X* como a variante de Palloni e Heligman.

23.4.3 Método do Filho Prévio

Cabe mencionar ainda o método do filho prévio, que se baseia num raciocínio semelhante aos métodos descritos acima, mas que exige dados por parte das mães no momento do seu nascimento do seu filho. Por isso, tipicamente, o método é aplicado em hospitais de maternidade. As perguntas feitas são extremamente simples, a saber:

- Esteve grávida antes ?
- O filho daquela gravidez nasceu vivo ?
- E está vivo atualmente ?
- Antes desse filho, teve outra gravidez ?
- O filho daquela gravidez anterior nasceu vivo ?
- E ainda está vivo atualmente ?

Eventualmente, é possível fazer apenas as primeiras três perguntas, sobre o filho prévio, e omitir as últimas três, sobre o filho penúltimo. Brass e Macrae (1984) estabeleceram que a proporção de filhos prévios que morreram corresponde aproximadamente a ${}_2q_0$ e a proporção de filhos penúltimos que morreram a ${}_3q_0$. Vaz Grave (1992) aplicou o método às mães que vinham dar à luz na Maternidade Lucrecia Paim em Luanda entre 1988 e 1991, encontrando um valor de 116 por mil para ${}_2q_0$ e 137 por mil para ${}_3q_0$.

23.5 MÉTODOS PARA ESTIMAR A MORTALIDADE ADULTA

23.5.1 Mães sobreviventes

A proporção de mães sobreviventes pode ser calculada de forma semelhante ao cálculo do número de filhos sobreviventes. Esta estimativa é feita de forma indireta quando perguntado a uma pessoa, no Censo Demográfico, se sua mãe está ou não viva naquele momento. Conforme Sawyer e Martín-Castilla (1989), a proporção de mães vivas de uma pessoa com idade x é o produto de dois fatores: o risco de sobreviver desde que tenha tido um filho e a exposição, que é o número de mulheres que tiveram filhos na época em que esta pessoa nasceu.

Os autores enumeram uma série de pressupostos que devem ser levados em consideração para a elaboração da estimativa, e Palloni, Massagli e Marcotte (1984) discutem o que ocorre com a análise quando alguns destes pressupostos são quebrados. Os pressupostos são:

1. Fecundidade constante;
2. Maternidade afetada pela estrutura etária e da fecundidade;
3. A mortalidade é alocada no tempo e linear no logito;
4. A população é fechada à migração;
5. A mortalidade não é diferencial por tamanho da família;
6. Não há processo de adoção; e
7. A mortalidade materna não é seletiva pela mortalidade dos filhos.

Apesar de interessante, segundo Brass (1975), uma série de problemas pode surgir quando da aplicação desta técnica. O primeiro problema prático é que os dados se referem a uma população selecionada, uma vez que não está refletindo a mortalidade das mulheres que nunca tiveram filhos. O segundo problema se refere ao cálculo da idade média da maternidade, pois quando ela é feita com base nas informações do último ano pode não ser representativa da distribuição de idade das mães no momento do nascimento dos filhos mais velhos. Um terceiro ponto está relacionado ao fato que uma mulher com muitos filhos será contabilizada muitas vezes. O quarto problema se refere às pessoas mais velhas terem tido um evento de mortalidade muitos anos atrás, o que pode provocar erros de memória. E, por último, o efeito da adoção, se esta for muito difundida na população pode provocar uma subestimação da mortalidade para as primeiras idades, pois as crianças adotadas reportarão que suas mães são vivas.

23.5.2 Métodos baseados na mensuração da cobertura do registro de óbitos

Para obter estimativas de mortalidade de forma direta, como descrito nos Capítulos 8 e 9, é preciso que as fontes de dados, de população e de óbitos, sejam confiáveis e de qualidade. No entanto, muitos países em desenvolvimento não apresentam qualidade dos dados satisfatória para que estas estimativas sejam realizadas diretamente. O registro civil é afetado por cobertura incompleta do registro de óbitos e por erros na declaração da idade dos falecidos. Uma alternativa seria a estimação das razões de sobrevivência da população por sexo e idade implícitas na comparação de dois censos sucessivos. Entretanto, esta comparação também pode apresentar vários problemas, tais como a cobertura diferencial (comparando entre idades e entre dois anos) e erros na declaração de idade (Naciones Unidas, 1986; Preston, Elo e Stewart, 1999). A pergunta sobre quantas pessoas faleceram no domicílio (agregado familiar) nos últimos 12 meses, por sexo, idade e eventualmente data do falecimento teoricamente compõe um registro de óbitos paralelo ao do registro civil para o ano anterior à data do censo. Entretanto, esta informação também está sujeita a vários erros potenciais:

1. Erro de referência de tempo, por exemplo a tendência a confundir anos calendário com anos corridos de 12 meses;
2. A tendência a evitar falar sobre acontecimentos desagradáveis como a morte de um familiar próximo;
3. A confusão sobre o significado exato de “neste domicílio” ou “que pertencia a este domicílio”, principalmente em casos onde a pessoa faleceu depois de ter passado um longo período no hospital;
4. A tendência a incluir pessoas que faziam parte da família, mas não do domicílio; e
5. Erros na declaração da idade da pessoa falecida.

Dadas essas limitações nas fontes de dados, várias metodologias foram desenvolvidas para que a mortalidade fosse estimada de forma indireta ou com base em modelos demográficos. Dentre os principais métodos para obter estimativas de mortalidade adulta, mais especificamente, e de idosos, podem-se apontar três grupos de metodologias normalmente aplicadas: 1) métodos que estimam a cobertura do registro de óbitos, relativa aos dados do censo (Brass, 1975; Preston et al., 1980; Bennett e Horiuchi, 1981; Hill, 1987; Hill, You e Choi, 2009; Moultrie et al., 2013; Hill, 2017); 2) métodos que se baseiam na sobrevivência de parentes (Brass, 1975; Naciones Unidas, 1986) e 3) métodos baseados na sobrevivência intercensitária (Brass, 1975; Naciones Unidas, 1986). Neste capítulo, o maior interesse está no primeiro grupo de métodos. Para o segundo grupo, o leitor é referido a Palloni, Massagli e Marcotte (1984) e a Moultrie et al. (2013: 222-256), entre outros.

As metodologias que serão aplicadas para estimar a mortalidade se referem à mensuração da cobertura do registro de óbitos (Brass, 1975; Preston et al., 1980; Bennett e Horiuchi, 1981; Hill, 1987; Hill, You e Choi, 2009; Moultrie et al., 2013; Hill, 2017), e à de sobrevivência intercensitária (Brass, 1975; Naciones Unidas, 1986). Os métodos usados na estimativa da cobertura do registro de mortes avaliam a cobertura do registro de óbitos comparando a distribuição etária de mortes com a distribuição etária da população (proveniente de censos ou inquéritos). Se a cobertura do registro de mortes relativa à cobertura do registro de pessoas pode ser estimada, os dados de mortalidade podem ser ajustados por esse diferencial de cobertura e um padrão de mortalidade não enviesado é calculado por idades em determinado período.

No conjunto de metodologias de estimativa da cobertura do registro de mortes, os métodos de distribuição de óbitos, todos eles têm um importante pressuposto na aplicação: “que a cobertura do registro de mortes não seja diferencial com a idade”, ou seja, admite-se que a estrutura de mortalidade seja correta e corrige-se apenas o nível da mortalidade. Os primeiros modelos desenvolvidos para estimar a cobertura do registro de óbitos, em relação aos dados de população (Brass, 1975; Preston et al., 1980) partem do forte pressuposto de população estável (constância nas estatísticas vitais ao longo do tempo). Os métodos posteriores (Bennett e Horiuchi, 1981; Hill, 1987; Hill, You e Choi, 2009; Moultrie, et al., 2013; Hill, 2017) eliminam esse pressuposto, mas mantêm a hipótese de população fechada ou sem diferencial migratório por idade.

23.5.3 Método da Equação de Balanço de Brass (1975) e de Preston et al. (1980)

O método da Equação de Balanço de Brass (1975) foi formulado a partir da relação de que a taxa de crescimento da população (ou de um segmento da população), (r_{x+}), deve ser igual à diferença entre a taxa de entrada (P_x/P_{x+}) e a de saída (D_{x+}/P_{x+}). Numa população fechada, a entrada na população de idade x e mais é dada pelas pessoas que completam exatamente x anos de idade (P_x) e as saídas consistem nas mortes das pessoas de x anos e mais (D_{x+}). Assim tem-se:

$$\frac{P_x}{P_{x+}} = r_{x+} + \frac{D_{x+}}{P_{x+}} \quad (23.25)$$

onde P_{x+} é o número total de pessoas acima da idade x . Numa população fechada e estável, provou-se que a relação acima é satisfeita. Numa população estável a taxa de crescimento da população a partir da idade x (r_{x+}) é constante para qualquer segmento de idade. No entanto, na

aplicação do método não é necessário o conhecimento prévio desta taxa. Seguindo a equação (23.25), nota-se que as taxas de entrada e de saída estão linearmente relacionadas. Ao considerar que do número total de mortes a partir da idade x (D_{x+}), somente uma proporção é registrada (D_{x+}^*), tem-se:

$$D_{x+}^* = C_x D_{x+} \quad (23.26)$$

C_x é um fator que representa a cobertura dos registros de morte para idades de x anos e mais. Como se supõe que o grau de cobertura no registro de mortes é o mesmo para qualquer grupo etário, ao menos acima da idade 5 ou 10, C_x pode ser substituído por uma constante C , que não varia com a idade. Desta forma, a equação (23.25) passa a ser:

$$\frac{P_x}{P_{x+}} = r + \frac{1}{C} \frac{D_{x+}^*}{P_{x+}} \quad (23.27)$$

Como a taxa de entrada é calculada a partir da distribuição etária da população, qualquer erro de cobertura não diferencial por idade é cancelado. Por outro lado, a taxa de morte calculada, utilizando o registro de mortes e a distribuição etária da população, é afetada por qualquer diferencial de cobertura entre o registro de mortes e de pessoas. Desta forma, para uma população fechada e estável, em que a cobertura do registro de mortes é constante para qualquer idade e a declaração de idade é confiável, a equação acima resulta num método pelo qual se pode estimar a cobertura do registro de mortes. A declividade da linha que relaciona as taxas de entrada e de saída estima o fator de correção do registro de mortes.

Os pressupostos básicos para a aplicação deste método são:

1. População estável (taxas de fecundidade e mortalidade constantes);
2. População fechada;
3. Grau de cobertura de mortes constante para as diferentes idades (ou seja, admite como verdadeira a estrutura de mortalidade observada);
4. Dados de dois censos de igual confiabilidade; e
5. Erros na declaração de idade não enviesados.

Geralmente uma boa forma de avaliar a quebra de pressupostos e possíveis ajustes é a partir de uma análise gráfica dos resultados. Normalmente, para contornar o não cumprimento de alguns pressupostos (como queda da fecundidade, população aberta à migração ou erro de declaração em idades avançadas) existe a opção de restringir o cálculo dos indicadores de cobertura para idades que sejam menos afetadas por esses fatores.

O método de Brass (1975) é mais robusto aos efeitos gerados pelo aumento na declaração da idade de óbito ou da idade da população (pressuposto 5), mas é mais vulnerável em casos de rápida queda da mortalidade e também requer muita atenção quando a fecundidade está declinando (quebra do Pressuposto 1).

O método de Preston et al. (1980) se baseia numa ideia relativamente direta. Ao considerar um número de pessoas numa determinada idade x numa população qualquer no tempo t , caso essas pessoas sejam acompanhadas até a morte, pode-se dizer que o número total de mortes (que ocorrem a partir da idade x) desse grupo de pessoas, num momento futuro, será igual à quantidade inicial de indivíduos que tinham idade x no momento t . Quando a população é estável e fechada, este número de mortes pode ser estimado pelo número corrente de mortes registradas a partir da idade x .

Como o número de óbitos acima da idade x , que ocorre depois do momento t , é estimado a partir da quantidade de mortes observadas num determinado ano, ao invés de acompanhar todas as pessoas até a morte, pode-se obter uma estimativa da cobertura do registro de mortes utilizando o número corrente de óbitos. Tendo P_x como o número de pessoas na idade exata x numa população estável com taxa de crescimento r , e D_x é o número de mortes à idade x , a estimativa do número de pessoas na idade x é dada por:

$$P_x = \sum_{y=x}^{\omega} D_y e^{r(y-x)} \quad (23.28)$$

Caso a população seja estável e fechada à migração, a taxa de crescimento é constante e vai estar corretamente especificada; e se as mortes são corretamente registradas o número observado de pessoas na idade x ($P(x)$) será igual ao número estimado de pessoas na idade x ($P'(x)$).

Caso o registro de óbitos não seja completo, a estimativa do sub-registro de óbitos é dada pela razão entre o número estimado de indivíduos na idade x (P'_x) e o número observado de pessoas na idade x (P_x). Para obter a medida da cobertura do registro de óbitos não se utiliza uma idade específica, e sim uma média de diferentes idades. Uma medida mais robusta, para minimizar os erros de declaração de idade, é a razão entre o número de pessoas estimado e o observado, calculados em grupos quinquenais de idade, ao invés de estimados por idade exata. Outra medida usada para anular distorções, resultantes de erros na declaração de idade e de cobertura não diferencial por idade, é considerar a razão entre o número estimado e observado de indivíduos, a partir de uma idade x até uma idade anterior ao grupo etário de intervalo aberto. O módulo PRECOA do pacote PASEX implementa o método automaticamente.

Os pressupostos para aplicação deste método são os mesmos daqueles mencionados no método de Brass (1975). No método de Preston et al. (1980) é difícil distinguir as diferentes violações dos pressupostos⁵ pela análise gráfica dos resultados. Ainda assim, a aplicação deste método se mostra mais robusta em relação à quebra do pressuposto de estabilidade (Pressuposto 1), permitindo que os resultados sejam avaliados e uma taxa de crescimento mais adequada seja utilizada. Por outro lado, conforme Paes (1993), este método é mais sensível a erros na declaração de idade (Pressuposto 5).

⁵ Por exemplo, a escolha de uma taxa de crescimento muito baixa, e tanto a variação da cobertura do registro de morte aumentando com a idade ou o aumento da idade na declaração de óbito produzem gráficos em que a sequência de $P'(x+)/P(x+)$ é crescente.

Apesar das técnicas acima mencionadas serem mais antigas, cada uma delas tem vantagens e limitações em sua aplicação, com relação à quebra dos pressupostos. Estudos recentes da mortalidade para o Brasil utilizam estas metodologias para obter estimativas de mortalidade, nas últimas décadas, a nível nacional (Paes, 1993; Oliveira, Albuquerque e Lins, 2004), de regiões (Gadelha et al. 2002; Campos e Rodrigues, 2004), de estados (Paes, 1993; Paes e Albuquerque, 1999; Paes, 2005) e cidades (Paes, 1993). Os estudos de Paes e Albuquerque (1999) e Paes (2005) utilizam os dois métodos acima mencionados para estimar a mortalidade, desagregada por Unidades da Federação, nos anos de 1980/1990 e 2000, respectivamente. Os resultados apontam para uma melhora na cobertura do registro de óbitos ao longo do tempo, e mais recentemente para as regiões do Norte e Nordeste do país, particularmente (Paes, 2005). Entretanto, o pressuposto de população estável é muito forte e fácil de ser violado. No caso da população brasileira e de outros países em desenvolvimento, em períodos mais recentes, houve avanços nos métodos buscando, principalmente, a aplicação de uma metodologia própria a populações não estáveis.

23.5.4 Método da Equação Geral de Balanço e Gerações Extintas combinado

O método da Equação Geral de Balanço (“Generalized Growth Balance” ou GGB, em inglês), formulado por Hill (1987), é uma generalização do método da Equação de Balanço de Brass (1975) e é aplicado para populações não estáveis, mas mantém os outros pressupostos⁶. Uma descrição mais atualizada do método encontra-se em Dorrington (2013 a). O método GGB é derivado da equação básica de equilíbrio demográfico, que define a taxa de crescimento da população como a diferença entre a taxa de entrada e a taxa de saída da população. Quando se têm disponíveis dois ou mais censos, a taxa de crescimento para cada segmento etário pode ser calculada, e o pressuposto de estabilidade não é mais necessário. Neste método, estima-se simultaneamente a cobertura do registro de mortes e a cobertura relativa da enumeração de dois censos.

Nesse método, assume-se que a relação entre as taxas de entrada e saída da população também ocorre para qualquer segmento de idade com intervalo aberto $x+$ (pessoas com x anos e mais de idade). Ou seja, numa população sem migração as entradas ocorrem como aniversários nas idades x . Dessa forma, a diferença entre a taxa de entrada em $x+$ e a taxa de crescimento populacional em $x+$ produz uma estimativa residual da taxa de mortalidade em $x+$. Intuitivamente, é possível avaliar que a taxa de entrada na população de 15 anos e mais, calculada pela taxa de pessoas fazendo 15 anos dividida pela população de 15 anos e mais, menos a taxa de mortalidade das pessoas de 15 anos e mais deve ser igual à taxa de crescimento da população de 15 anos e mais. A diferença calculada entre a taxa de entrada nesse segmento de idade e a taxa de crescimento desse segmento fornece uma estimativa da taxa de mortalidade do segmento 15 anos e mais, assumindo que não há fluxos migratórios significativos. Essa estimativa residual de mortalidade pode ser comparada com a taxa de mortalidade do mesmo segmento etário, calculada a partir de dados censitários ou de registros vitais, e as diferenças entre eles podem ser analisadas como erros na declaração do registro de óbitos da declaração do tamanho da população.

Se a estimativa residual de mortalidade puder ser estabelecida a partir de dois censos populacionais, e comparada com uma estimativa direta de mortalidade, usando o registro de óbitos

⁶ Bhat (2002) formulou uma versão da técnica que, além de relaxar o suposto da estabilidade, não exige que a população esteja fechada à migração.

ou enumeração de óbitos do censo demográfico, o grau de cobertura do registro de óbitos pode ser estimado relacionando esses dois movimentos. Noutras palavras, a partir da relação entre a diferença da taxa de entrada e a taxa de crescimento com a taxa de mortalidade, em cada grupo de idade, é possível estimar um intercepto, que captura qualquer variação na cobertura entre os dois censos/inquéritos, como também é possível estimar uma inclinação que serve como indicador do grau de cobertura do registro de mortes, em relação à média da cobertura de ambos os censos/inquéritos. Deve-se notar, que o método compara a distribuição etária dos óbitos (média no período intercensitário) com a mudança populacional entre os censos. Especificamente, a estimativa se refere à cobertura do registro entre os censos, e não a uma data específica.

Assim, a equação (23.25) pode ser reescrita como:

$$\frac{P_x}{P_{x+}} - r_{x+} = \frac{D_{x+}}{P_{x+}} \quad (23.29)$$

Considerando que k_1 e k_2 representam a cobertura da enumeração do primeiro e segundo censos, respectivamente, e C é um fator que representa a cobertura dos registros de morte, tem-se a seguinte equação:

$$P_1^*(x+) = k_1 \cdot P_1(x+); P_2^*(x+) = k_2 \cdot P_2(x+); D^*(x+) = C \cdot D(x+) \quad (23.30)$$

onde $P_1^*(x+)$ e $P_1(x+)$ correspondem, respectivamente, ao número de pessoas de idade x e mais observadas (enumeradas) no primeiro censo e o número real; $P_2^*(x+)$ e $P_2(x+)$ correspondem ao número de pessoas de idade x e mais observadas (enumeradas) no segundo censo e o número real de pessoas, respectivamente; $D^*(x+)$ e $D(x+)$ correspondem, respectivamente, ao número de mortes pessoas de idade x e mais observadas no registro de mortes e o número real de mortes.

Substituindo e rearranjando as fórmulas acima tem-se que:

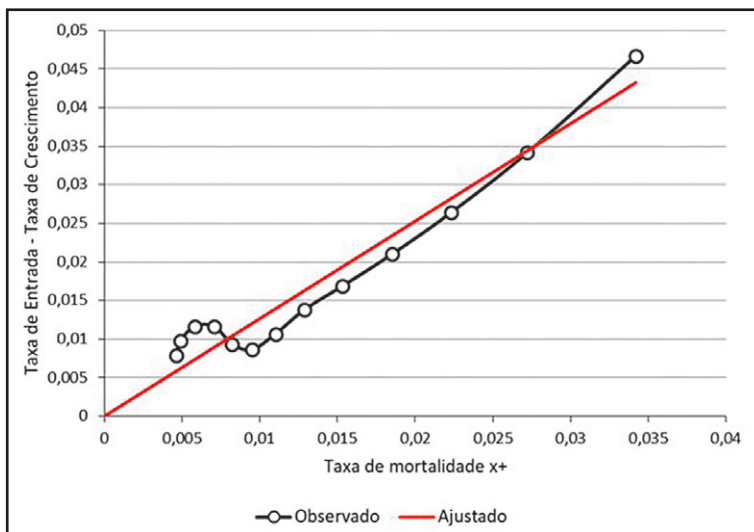
$$\frac{P^*(x)}{P^*(x+)} - r^*(x) = \frac{1}{t} \ln \left(\frac{k_1}{k_2} \right) + \frac{(k_1 \times k_2)^{1/2}}{C} \left(\frac{D^*(x+)}{P^*(x+)} \right) \quad (23.31)$$

Desta forma, a taxa de entrada menos a taxa de crescimento tem uma relação linear com a taxa de mortes, podendo-se calcular a partir dessa equação a cobertura do registro de mortes (C) e a cobertura relativa da enumeração de dois censos (k_1/k_2). Lembrando que, os pressupostos para aplicação do método da Equação Geral do Balanço são os mesmos daqueles do método que lhe deu origem, eliminando apenas o de população estável. Os valores de $P^*(x)$ e $P^*(x+)$ em (23.31) usualmente são calculados em termos das médias geométricas de $P_1^*(x)$, $P_2^*(x)$, $P_1^*(x+)$ e $P_2^*(x+)$.

O módulo CENCT de MORTPAK aplica (23.28) para estimar a cobertura de um censo em relação ao outro, enquanto o módulo BENHR faz a análise de Bennett-Horiuchi automaticamente, mas a lógica subjacente de ambos os procedimentos é relativamente fácil de entender graficamente. No eixo X estão as idades observadas e o eixo Y representa a estimativa de cobertura para cada um dos intervalos etários. A estimativa do grau de cobertura é obtida a partir de uma regressão ortogonal nos pontos para os segmentos etários considerados. A inclinação da reta estima o fator de ajuste necessário para corrigir as taxas de mortalidade observadas. O intercepto da reta estimada

fornece uma estimativa da cobertura relativa entre os dois censos utilizados na análise. Em geral, recomenda-se não considerar as idades abaixo de 30 anos para estimar o sub-registro. O Gráfico 23.6 mostra o exemplo do Maranhão para o período de 2000 a 2010.

Gráfico 23.6: Gráfico de diagnóstico do Método Equação Geral do Balanço, homens, Maranhão, 2000-2010



Fonte: Queiroz et al. (2017).

A análise do gráfico de dispersão confirma a preocupação com a suposição de população fechada nessas unidades federativas. Os pontos em idades mais jovens, principalmente para os homens, apresentam maior distanciamento da reta estimada. A forma do gráfico parece indicar diversos problemas: fluxos migratórios, declínio dos níveis de mortalidade, variação de cobertura dos censos demográficos e diferencial no registro de óbitos.

O método das Gerações Extintas Sintéticas (“Synthetic Extinct Generations” ou SEG, em inglês), desenvolvido por Bennett e Horiuchi (1981), parte do método de Preston et al. (1980). O método se baseia no conceito de estabilidade generalizada (ver seção 22.6 do Capítulo 22) e as suas implicações para o crescimento das populações por faixa etária. Uma descrição mais recente encontra-se em Dorrington (2013 b). O método é utilizado quando a população desvia da estabilidade, ou seja, a taxa de crescimento r varia com a idade. Intuitivamente, o método mostra que a população de uma determinada idade hoje deve ser igual a todas as mortes que ocorrerão a essa população no futuro. Mortes atuais acima de determinada idade podem ser ajustadas pela taxa de crescimento específica daquele grupo etário flexibilizando a ideia de população estável, segundo as linhas da população estável generalizada discutidas na seção 22.6 do Capítulo 22.

Nestes casos, a estimativa do número de pessoas na idade $x-5$ é dada por:

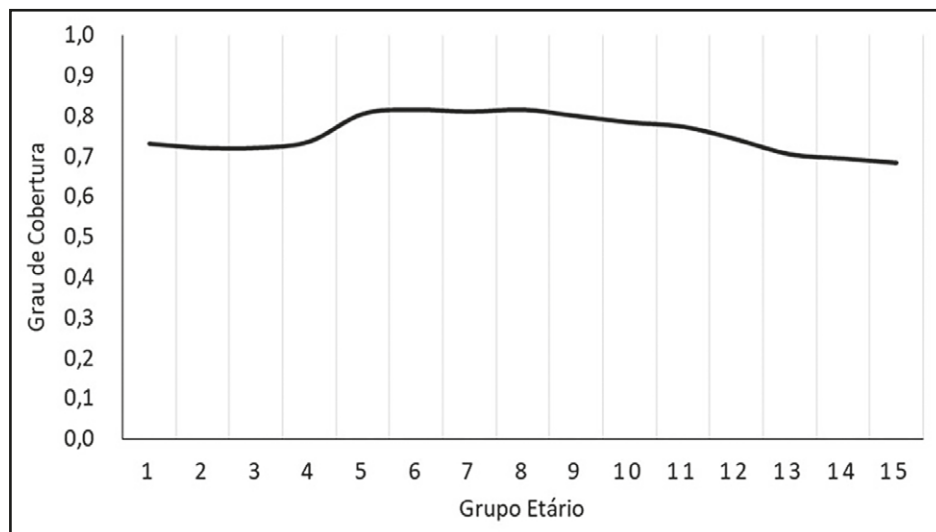
$$P_{x-5} = P_x e^{5r_{x-5}} + {}_5D_{x-5} e^{2,5r_{x-5}} \tag{23.32}$$

Neste método, a estimativa do sub-registro de óbitos também é dada pela razão entre o número estimado de pessoas na idade x ($P'(x)$) e o número observado de pessoas na idade x

($P(x)$). Na aplicação do método de Bennett e Horiuchi (1981) os pressupostos são os mesmos de Preston et al. (1980), exceto o de estabilidade populacional. Ou seja, o método usa taxas de crescimento específicas por idade para converter uma distribuição de óbitos por idade numa distribuição etária de população. Uma vez que numa população os óbitos observados a partir de uma determinada idade x são iguais à população da idade x , ajustada pela taxa de crescimento populacional por intervalo etário, os óbitos de uma população na idade $x+$ fornecem uma estimativa da população daquela idade x . O grau de cobertura do registro de óbitos será dado, então, pela razão entre os óbitos estimados pela população acima da idade x e a população observada acima da idade x .

O Gráfico 23.7 mostra a relação entre a população estimada via óbitos e a população observada. A razão entre as duas grandezas fornece a estimativa do grau de cobertura do registro de óbitos. Há uma longa discussão sobre qual intervalo etário deve ser usado para se obter essa estimativa. Hill (2017) sugere o uso do segmento etário acima de 35 anos para evitar possíveis efeitos de fluxos migratórios. Todavia, deve-se analisar também o perfil da curva e evitar o uso de pontos de grande flutuação.

Gráfico 23.7: Relação entre a população estimada via óbitos e a população observada, homens, Maranhão, 2000-2010



Fonte: Queiroz et al. (2017).

Em resumo, os métodos da Equação Geral de Balanço (Hill, 1987) e da Geração Extinta (Bennett e Horiuchi, 1981), relativamente mais recentes, têm a vantagem de eliminar o pressuposto da estabilidade da população. A aplicação do método da Equação Geral de Balanço permite calcular, além do sub-registro intercensitário de óbitos, o diferencial de cobertura entre dois censos, eliminando assim tanto o pressuposto de população estável quanto o problema de diferenças de confiabilidade entre os dois censos.

Assim, o método Geração Extinta estima a cobertura do registro de óbitos e pode ser conjugado com a correção do diferencial de cobertura de dois censos, sendo este conhecido como método Geração Extinta Combinado. Este método pode apresentar melhores resultados nos

casos em que se verifica um diferencial de cobertura entre dois censos (Hill, 2003; Hill, You e Choi, 2009; Hill, 2017).

Hill, You e Choi (2009) afirmam que, de forma geral, o método Geração Extinta Combinado se mostrou menos sensível a erros na declaração de idade, da população ou de morte, ou ao diferencial de cobertura em comparação com o método da Equação Geral de Balanço. Por outro lado, o método Geração Extinta Combinado é bastante afetado pela migração, enquanto o da Equação Geral de Balanço se mostrou menos sensível à quebra desse pressuposto. Murray et.al (2010) avaliam a estimativa de cobertura dos métodos de distribuição de mortes usando uma série de simulações, num total de 234 variações dos métodos e 78 diferentes segmentos de idade. Os autores avaliam, a partir dessa análise, três variantes dos métodos que melhor se adequam os diferentes cenários usados. Eles concluem que usar o método de Bennett-Horiuchi (SEG) com segmento 55-70 anos, a generalização do GGB com segmento 40-70 e a combinação dos métodos com segmentos de idade 50-70 apresentam as melhores desempenhos. Hill, You e Choi (2009) recomendam a utilização do segmento 5-65 anos ao invés de um intervalo mais curto iniciando numa idade mais avançada. Mais recentemente, Hill (2017) sugere o uso de um segmento mais amplo de idade para aplicação do método GGB (5-65) combinado com um segmento mais curto para as gerações extintas (SEG) de 50-70 anos. O argumento de Hill é que o GGB tem uma boa capacidade de estimar a qualidade de um censo em relação ao outro e permitiria o melhor ajuste do método.

Uma opção alternativa é proposta no pacote DDM (de “Death Distribution Methods”) do R-CRAN que faz uma busca, por meio de métodos estatísticos. A ideia principal é construir um estimador que forneça a melhor solução de todas as combinações possíveis de faixas etárias. O estimador implementado pelo pacote DDM é o que minimiza a Raiz do Erro Quadrático Médio (“Root Mean Square Error” ou RMSE, em inglês). A Raiz do Erro Quadrático Médio é uma medida da diferença entre os valores preditos por um modelo e os valores realmente observados. O segmento de idade que minimiza o RMSE é considerado o segmento ideal da estimativa de cobertura do GGB. Noutras palavras, a solução automática é determinar a cobertura do registro de óbitos com base no segmento etário que produz a menor diferença residual entre as taxas de mortalidade observadas e estimadas e sua linha de ajuste.

Os métodos mais recentes de distribuição de óbitos têm limitações em relação ao pressuposto de população fechada. Teoricamente, os fluxos migratórios podem ser facilmente incorporados aos métodos (Hill e Queiroz, 2010), mas em geral estimativas de migração por idade são menos comuns e de pior qualidade do que estimativas de mortalidade adulta. Hill e Queiroz (2010) buscam ajustar esta metodologia incorporando a migração quando não se tem o perfil de migração para a população em estudo. A aplicação desse método ajustado é feita em dois passos e de forma iterativa. Primeiro, estima-se taxas líquidas de migração específicas por idade por meio de um modelo padrão de migração e depois, ajusta-se o método da Equação Geral de Balanço para esta migração. As estimativas sugerem que o procedimento proposto apresenta bons resultados em populações com dados de boa qualidade e com altas taxas de migração líquida.

Uma questão que não é tratada por nenhum dos métodos aqui expostos é a de que o grau de cobertura de mortes seja constante para as diferentes idades. Oliveira, Albuquerque e Lins (2004) pressupõem que a cobertura dos óbitos nas idades mais avançadas seja diferenciada da cobertura nas idades adultas. Nesse sentido, os autores apresentam um fator de redução na estimativa de sub-registro calculada que foi aplicado para o Brasil, por grupo de idade quinquenal (a partir dos

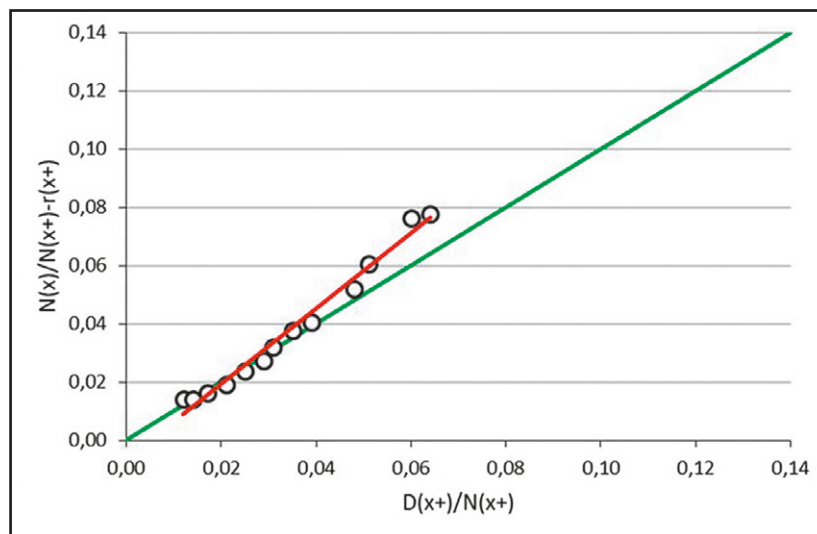
35 anos, grupo aberto 80 anos e mais). No entanto, grande parte dos estudos de mortalidade para o Brasil, inclusive os que tratam especificamente da mortalidade dos idosos, utilizam nas idades adultas e nas mais avançadas o mesmo grau de correção da cobertura do registro de óbitos (Paes, 1993; Paes e Albuquerque, 1999; Paes, 2005; Campos e Rodrigues, 2004), sendo esta ainda uma limitação destes métodos.

23.5.5 Aplicação dos métodos GGB e SEG

Os Censos Demográficos de 1997 e 2007 em Moçambique oferecem uma oportunidade interessante para a aplicação dos métodos de correção de sub-registro e estimativas de mortalidade em contexto com ausências de um sistema de registro civil adequado. Nos dois Censos, foi incluída a pergunta sobre óbitos no domicílio nos 12 meses anteriores à realização do Censo. Caso tenham ocorrido óbito(s) no domicílio, o respondente informa o sexo e a idade do falecido. No caso de Moçambique, caso o óbito seja feminino ainda se obtém informação para gerar estimativas de mortalidade materna.

Os resultados apresentados aqui foram obtidos do artigo de Alberto e Queiroz (2015) e também desenvolvidos por Alberto (2013). O Gráfico 23.8 mostra os resultados para os homens do GGB (Hill, 1987) para os homens no período 1997-2007. O eixo X mostra a taxa de mortalidade observada para as idades $x+$, e o eixo Y representa as taxas de mortalidade para as idades $x+$, derivada como resíduo das taxas de crescimento e entrada nas idades $x+$.

Gráfico 23.8: Gráfico de Diagnóstico, Equação de Balanceamento (GGB), Homens, Moçambique 1997-2007

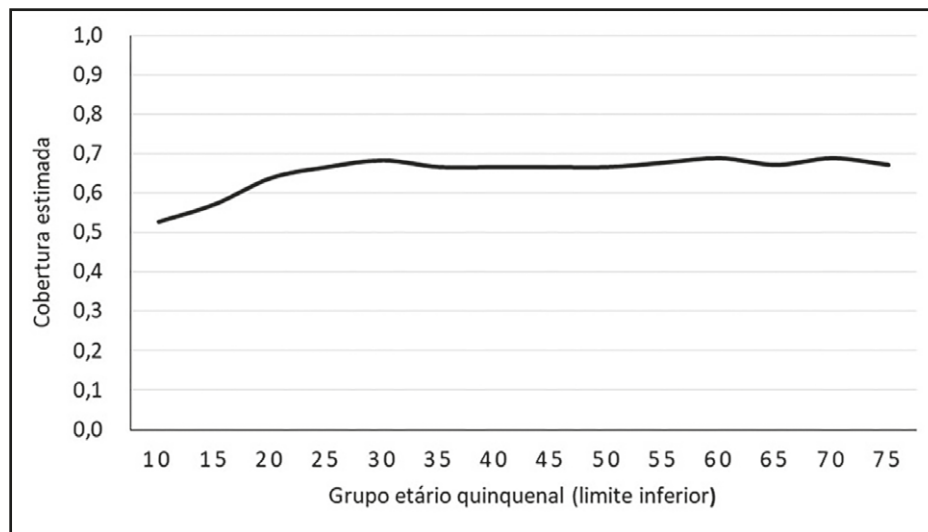


Fonte: Alberto e Queiroz (2015).

O Gráfico 23.9 mostra os resultados do método de Bennett-Horiuchi para o mesmo período em Moçambique. No eixo X estão as idades observadas e o eixo Y representa a estimativa de cobertura para cada um dos intervalos etários. Os resultados ajudam a definir qual intervalo etário utilizar para obter as estimativas de fator de correção. Os resultados são bastante estáveis a partir da idade 30 e mais e não indicam nenhum problema adicional com os dados. Todavia, considerar

as idades abaixo de 30 anos para estimar o sub-registro não é recomendável. A forma do gráfico parece indicar diversos problemas: fluxos migratórios, declínio dos níveis de mortalidade, variação de cobertura dos censos demográficos e diferencial no registro de óbitos. O principal resultado é que os dados indicam um alto grau de subenumeração de óbitos em Moçambique.

Gráfico 23.9: Gráfico de Diagnóstico – Método de Gerações Extintas (SEG), Homens, Moçambique, 1997-2007



Fonte: Queiroz et al. (2017).

Alberto e Queiroz (2015) mostram que as estimativas apresentam pouca variação independente do método utilizado. Todavia, eles mostram diferentes testes de estimativas usando intervalos de idade diferentes para obter o ajuste das curvas. No período entre 1997 e 2007, as coberturas de óbitos oscilaram entre 0,658 e 0,908 entre as diferentes regiões e os diferentes grupos etários usados nas estimativas. O menor valor de cobertura de óbitos foi encontrado no método no grupo de ajuste de 25-59 anos e o maior valor foi achado no método de gerações extintas quando foi usado o grupo de ajuste de 15-59 anos. De modo geral, os resultados indicam que o censo demográfico capta entre 65 e 90% dos óbitos ocorridos em Moçambique nos dozes meses anteriores à realização do censo, implicando que estimativas de mortalidade e esperança de vida com dados observados diretamente são subestimadas.

Com base nas estimativas de grau de cobertura da enumeração dos óbitos, é possível calcular as probabilidades de morte adulta (${}_5q_{15}$). No período analisado, os autores estimam que a mortalidade adulta masculina estaria ao redor de 0,60, ou seja, a probabilidade de uma pessoa de 15 anos morrer antes de completar 60 anos era de 60%. Para as mulheres, a estimativa apresentada pelos autores foi de 0,55. Para o período entre 1997 e 2007, a maior probabilidade de morte masculina nas idades entre 15 e 60 anos verificada no país é explicada, para além de outros fatores, pela maior prevalência da AIDS (SIDA). A menor probabilidade observada nos dados diretos pode ser explicada pela subestimação das mortes no país resultante da omissão.