

MODELAGENS DA MORTALIDADE, FECUNDIDADE E MIGRAÇÃO

20.1 INTRODUÇÃO

Os métodos tradicionais usados na demografia não fazem supostos sobre o comportamento das funções de mortalidade, fecundidade e migração subjacentes. Por exemplo, o cálculo da esperança de vida de uma tábua de vida faz alguns supostos sobre a regularidade local de ℓ_x e $\mu(x)$, mas não supõe nada a priori sobre a forma mais global dessas funções, por exemplo alguma relação entre a mortalidade adulta e a mortalidade na infância. Entretanto, essas relações existem. O fato de que a mortalidade na infância de um determinado país é baixa geralmente dá uma boa indicação de que a mortalidade adulta também deve ser baixa. Essas regularidades permitem a formulação de modelos, principalmente modelos que descrevem o perfil etário dos fenômenos demográficos. O uso de modelos na demografia visa alguns objetivos diferentes que podem ser resumidos da seguinte forma:

1. Graduação: O objetivo aqui é apresentar fórmulas matemáticas teóricas representativas para o comportamento das variáveis demográficas como a mortalidade ou a fecundidade que possam ser comparadas com dados observados, para eliminar erros prováveis ou oscilações aleatórias nestes. Os parâmetros dessas fórmulas geralmente servem para otimizar o ajuste da fórmula a uma variedade de padrões empíricos observados no mundo e não necessariamente possuem uma interpretação teórica muito clara que permita a sua projeção para o futuro.

2. **Projeção:** Os modelos usados para fins de projeção geralmente são mais simples, com menos parâmetros, do que os modelos usados para a graduação. O seu objetivo primordial não é obter o melhor ajuste possível a dados empíricos existentes, mas captar as dimensões principais do fenômeno que permitam a sua interpretação e projeção para o futuro. Por exemplo, como será visto abaixo, no caso da fecundidade as dimensões principais podem ser concebidas como o número médio de filhos que uma mulher terá na sua vida, a idade média na qual ela terá os filhos e o número médio de anos entre o nascimento do primeiro e do último filho. Existem outras maneiras para parametrizar a fecundidade, mas a vantagem desta é a simplicidade conceitual dos parâmetros que facilita a sua projeção para o futuro.
3. **Estimação indireta dos fenômenos:** O próximo capítulo trata dos métodos indiretos que estimam a mortalidade, fecundidade ou migração a partir das suas consequências ou a partir de perguntas resumidas que geralmente não fornecem muito detalhe sobre padrões etários e variações no tempo. Portanto, esses métodos precisam usar informação sobre o comportamento esperado da mortalidade, fecundidade e migrações por idade para preencher as lacunas ou as deficiências nos dados observados. Da mesma forma como acontece na estatística, o conhecimento do comportamento das distribuições subjacentes dos fenômenos permite uma maior eficiência na estimação dos parâmetros e possibilita estimativas de algumas quantidades que de outra forma seriam pouco factíveis.

Tratamentos mais detalhados sobre o papel de modelos formais na demografia podem ser encontrados em Newell (1988), Coale e Trussell (1996) e em Burch (2018).

20.2 MODELOS DE MORTALIDADE

Os modelos de mortalidade por idade podem ser divididos em três categorias:

1. A primeira categoria são os modelos puramente matemáticos, baseados em curvas analíticas que descrevem funções da tábua de vida, pelo menos em alguma parte da escala etária, por exemplo na infância ou na velhice.
2. A segunda categoria são os *modelos relacionais*, que partem de uma tábua de vida empírica, mas transformam uma das suas funções. O caso mais simples é o sistema de tábuas de vida obtidas ao multiplicar a função $\mu(x)$ de uma tábua de vida existente por fatores de intensidade r .
3. Os modelos mais usados são os modelos sintéticos, baseados na junção de grandes números de tábuas de vida empíricas, para descrever a variedade das experiências em diferentes níveis de mortalidade historicamente observados.

20.2.1 Modelos puramente matemáticos

Esta é a categoria mais antiga. Hoje em dia ela é usada principalmente nos estudos atuariais que têm a vantagem de lidar principalmente com a população adulta, de modo que a parte mais

complexa da curva, até os 20 ou 30 anos, não precisa ser modelada. A formulação dos modelos geralmente parte da função $\mu(x)$, a força da mortalidade. Aqui serão discutidas algumas das fórmulas mais conhecidas; para informação adicional sobre o tema, pode-se consultar Mina-Valdés (2001) e Pitacco (2019), entre outros.

20.2.1.1 A Lei de Gompertz e suas extensões

Segundo Gompertz (1825), o risco de morte aumenta exponencialmente com a idade a uma taxa instantânea de mortalidade definida por:

$$\mu(x) = B \exp(c x) \quad (20.1)$$

onde B e c são parâmetros com sinais positivos e x é a idade. Por sua vez, o número de sobreviventes (ℓ_x) nesta especificação decresce por:

$$\ell(x) = \exp\left(\frac{B}{c} (1 - \exp(c x))\right) \quad (20.2)$$

Uma vez derivada a função $\ell(x)$ na relação (20.2), as outras funções da tábua de vida podem ser completadas usando as fórmulas usuais descritas no Capítulo 9.

Makeham (1860) introduziu na Lei de Gompertz um parâmetro A que busca incorporar a mortalidade por causas externas:

$$\mu(x) = A + B \exp(c x) \quad (20.3)$$

Já a função de sobreviventes neste caso tem a seguinte forma:

$$\ell(x) = \exp\left(-A x + \frac{B}{c} (1 - \exp(c x))\right) \quad (20.4)$$

Evidentemente, a curva descrita da forma (20.1) ou (20.3) não capta a mortalidade infantil e de crianças. Ambas as fórmulas foram concebidas com o propósito de descrever a mortalidade além dos 30 ou 40 anos de idade, para fins atuariais. A sua aplicação mais comum é para “terminar” uma tábua de vida além das idades para as quais se dispõe de dados observados. Um exemplo desse tipo de procedimento para terminar ou extrapolar uma tábua de vida foi descrito na seção 9.7 do Capítulo 9, mas nesse caso foi usada uma fórmula improvisada que garantisse a consistência com uma esperança de vida previamente estimada. Quando não existem limitações deste tipo, as fórmulas (20.2) ou eventualmente (20.4) geralmente oferecem uma alternativa melhor.

A outra limitação das fórmulas (20.1) e (20.3) é que elas exageram o nível de mortalidade nas idades mais avançadas, além dos 85 anos. Segundo investigações empíricas que vêm sendo desenvolvidas desde os anos 80, o nível de mortalidade não continua aumentando cada vez mais com a idade, mas tem uma tendência ao nivelamento (Horiuchi e Wilmoth, 1998). Consequentemente,

foram propostas especificações alternativas que supostamente expressam melhor qual é o perfil da mortalidade nessas idades. A mais conhecida é a especificação logística de Kannisto (1996; Thatcher, Kannisto e Vaupel, 1998), que modifica (2.4), acrescentando um termo no denominador:

$$\mu(x) = A + B \exp(c x) / (1 + B \exp(c x)) \quad (20.5)$$

Para valores elevados de x , (20.5) não cresce sem limites mas tende para $A+1$.

Outra especificação que evita o crescimento excessivo de $\mu(x)$ nas idades avançadas é aquela proposta por Coale e Kisker (1990) que é igual à função de Gompertz (20.1), com a diferença de que para idades superiores a 85 anos (20.1) se converte em

$$\mu(x) = B \exp(c x - \frac{1}{2} s (x - 85)^2) \quad \text{para } x > 85 \quad (20.6)$$

onde s é um parâmetro positivo que reduz o aumento da mortalidade nas idades avançadas.

É preciso assinalar que a noção de um limite no aumento do nível de mortalidade nas idades mais avançadas, embora maioritariamente aceita pela comunidade de estudiosos, não é totalmente consensual. Alguns atuários (Gavrilov et al., 2014) afirmam que a tendência aparente de nivelamento se deve a problemas com os dados sobre a mortalidade nas idades extremas e que a verdadeira tendência se aproxima mais a (20.1) ou (20.3) do que a (20.5). Por exemplo, se existe uma tendência sistemática ao exagero das idades mais elevadas, o resultado será que a mortalidade em cada idade observada parecerá mais baixa do que realmente é.

20.2.1.2 Fórmulas que reproduzem a mortalidade em todas as idades

Como se notou acima, uma das limitações da curva de Gompertz e suas variantes é que elas não reproduzem adequadamente qual é o comportamento da mortalidade nas idades mais baixas, até os 15 ou 30 anos. Existem algumas especificações alternativas que resolvem esse problema ao somarem componentes distintos da mortalidade, onde cada uma representa a mortalidade num determinado grupo etário. Uma dessas especificações é o modelo de Thiele (1874), que tem a seguinte forma:

$$\mu(x) = B_1 \exp(-c_1 x) + B_2 \exp(-\frac{1}{2} c_2 (x-m)^2) + B_3 \exp(c_3 x) \quad (20.7)$$

Embora a fórmula seja pouco usada hoje em dia, ela ilustra bem qual é a ideia atrás das especificações híbridas desse tipo. O último termo é uma função convencional de Gompertz, que capta a mortalidade nas idades mais avançadas. O primeiro termo descreve a mortalidade na infância com uma função parecida com a de Gompertz, mas usando um expoente negativo, para que seja decrescente com a idade. O termo do meio, finalmente, que tem a forma de uma distribuição normal, descreve a elevação da mortalidade, particularmente da mortalidade masculina, aproximadamente entre as idades de 15 e 35, devido a acidentes e outras causas externas nesse grupo de

idade. Nos países onde a mortalidade materna é elevada, o padrão etário da mortalidade também se caracteriza por um aumento nessa faixa. Finalmente, nos países africanos de alta incidência de AIDS (SIDA), esta também contribui à elevação da mortalidade nas mesmas idades.

Quando se substitui um constante no lugar do termo central de (20.7), o resultado é uma expressão mais simples conhecida como o modelo de Siler (1979) que em anos recentes tem atraído um certo interesse por parte dos especialistas em mortalidade (Cohen, Bohk e Rau, 2018; Engelman, Caswell e Agree, 2014; Goldstein e Wachter, 2006). Gage e Mode (1993) argumentam que, apesar da sua simplicidade, o modelo fornece um ajuste melhor a muitos padrões empíricos de mortalidade do que o modelo mais complexo de Heligman e Pollard, que será apresentado abaixo. A forma geral da curva proposta por Siler é a seguinte:

$$\mu(x) = B_1 \exp(-c_1 x) + B_2 + B_3 \exp(c_3 x) \quad (20.8)$$

Hoje em dia se usa mais a curva de Heligman e Pollard (1980), que é muito parecida com (20.7), mas com termos um pouco distintos:

$$\frac{q_x}{1-q_x} = A^{(x+B)^C} + D e^{-\frac{1}{2} E (\ln(x)-\ln(F))^2} + GH^x \quad (20.9.a)$$

ou alternativamente

$$q_x = A^{(x+B)^C} + D e^{-\frac{1}{2} E (\ln(x)-\ln(F))^2} + \frac{G H^x}{1 + G H^x} \quad (20.9.b)$$

O terceiro termo dessa expressão, que aqui foi escrito com os nomes dos parâmetros geralmente usados para descrever a curva de Heligman-Pollard, é simplesmente uma função de Gompertz. Kostaki (1992) sugere acrescentar um parâmetro K , substituindo o x no terceiro termo por Kx , o que transforma (20.9.a-b) num modelo de 9, em vez de 8, parâmetros. A curva especificada em (20.9.a-b) se ajusta bem a uma variedade de tábuas de vida empíricas, principalmente se for aplicada com o parâmetro adicional. Isso a faz ideal para fins de graduação.

Os problemas da curva (20.9.a-b) são, por um lado, que é relativamente difícil ajustar uma curva tão complexa aos dados, não só devido ao número de parâmetros mas também porque os parâmetros são altamente correlacionados. Os programas MORTPAK (ver seção 17.2) e MortalityLaws (ver seção 17.1), da biblioteca de aplicativos em “R” mantida pelo CRAN, que além das especificações de Heligman-Pollard, Thiele, Kannisto, Makeham e Gompertz permite ajustar uma variedade de outros modelos de mortalidade, resolvem esse problema. Por outro lado, mesmo que seja possível estimar todos os parâmetros para a graduação de dados empíricos, não há uma interpretação intuitiva de todos eles que permita, por exemplo, fazer projeções baseadas nessa especificação. As especificações discutidas a seguir, mesmo que não se ajustem tão bem aos dados quanto (20.8), têm a vantagem de uma estrutura mais simples que facilita a interpretação dos seus parâmetros e o seu uso em projeções da mortalidade.

20.2.2 Modelos relacionais

20.2.2.1 Mortalidade proporcional

O modelo mais simples deste tipo já foi introduzido em (17.48) do Capítulo 17. Trata-se do chamado modelo proporcional. Suponha que o nível de mortalidade em todas as idades seja multiplicado por um fator r , ou seja, $\mu_r(x) = r \mu(x)$. Então a função $\ell_r(x)$ correspondente a este novo nível de mortalidade seria

$$\ell_r(x) = \exp\left(-\int_0^x r \mu(t) dt\right) = \ell^r(x) \quad (20.10)$$

ou seja, multiplicar toda a função por um fator r equivale a elevar a função $\ell(x)$ ao expoente r . As outras funções da tábua de vida podem ser derivadas facilmente de (20.10). Quando $r > 1$, a esperança de vida da tábua transformada será menor do que na tábua original e vice-versa. Como se viu no Capítulo 17, o tamanho da mudança está associado ao conceito de *entropia* da tábua de vida original.

O modelo proporcional é relativamente pouco usado na demografia, mas tem uma aplicação importante na bioestatística, nos chamados modelos de Cox (1972), que são usados para determinar como diferentes fatores contribuem ao nível de mortalidade de diferentes indivíduos com um padrão subjacente igual de mortalidade por idade. A discussão desse tipo de modelo de análise está além do escopo deste livro.

20.2.2.2 O modelo baseado em logitos

Brass (1974 a) estabeleceu a seguinte relação aproximada entre as funções ℓ_x de diferentes tábuas de vida:

$$\frac{1}{2} \ln\left(\frac{1}{\ell_x^{(1)}} - 1\right) = A + B \frac{1}{2} \ln\left(\frac{1}{\ell_x^{(2)}} - 1\right) \quad (20.11)$$

A função $\frac{1}{2} \ln(1/x - 1)$ é conhecida como *logito*(x). O fator $\frac{1}{2}$ não é estritamente necessário, mas é geralmente acrescentado, por razões históricas. Portanto, (20.11) pode ser escrito também como

$$\text{logito}\left(\ell_x^{(1)}\right) = A + B \text{logito}\left(\ell_x^{(2)}\right) \quad (20.12)$$

Em vez de estabelecer a relação (20.12) para tábuas de vida arbitrárias, é mais conveniente sempre escolher a mesma tábua padrão para o ℓ_x da direita, ou seja,

$$Y(x) = \text{logito}(\ell_x) = \alpha + \beta \text{logito}(\ell_x^p) = \alpha + \beta Y_p(x) \quad (20.13)^1$$

¹ É de notar que a principal utilidade de (20.13) é a conveniência de representar a relação em termos lineares, o que facilita o ajuste. Mas a função ℓ_x também pode ser representada diretamente, sem recorrer à transformação logito, pela seguinte expressão: $\ell_x = (\ell_x^p)^\beta / ((\ell_x^p)^\beta + e^{2\alpha}(1 - \ell_x^p)^\beta)$.

A função padrão $Y_p(x)$ tem sido tabulada para um padrão conhecido como o Padrão Geral de Brass (“Brass General Standard”, em inglês), o que permite caracterizar uma função de sobrevivência ℓ_x – e portanto uma tábua de vida – inteira apenas em termos dos seus parâmetros α e β relativos a esse padrão. O parâmetro α é um indicador do nível geral de mortalidade: quanto maior for α , menor será a esperança de vida. A interpretação do parâmetro β é menos óbvia. Valores de β acima de 1 aproximam ℓ_x a 1 para as idades x menores e aproximam ℓ_x a 0 nas idades mais avançadas. O resultado é uma mortalidade menor na infância e na velhice, enquanto aumenta a mortalidade nas idades intermediárias.

Os valores de $Y_p(x)$ para $x=1$ até 99 encontram-se no Apêndice 3. O padrão geral de Brass não é o único sendo usado como referência para construir famílias de modelos. Na África, Brass (1968) sugere que, em vez do Apêndice 3, é melhor usar um padrão que se adapta melhor às condições africanas de mortalidade. Esse chamado *padrão africano* de mortalidade se caracteriza por uma mortalidade relativamente alta de crianças de 1-4 anos em comparação com a mortalidade infantil. As únicas diferenças do padrão africano são os valores de $Y_p(1), \dots, Y_p(9)$; os demais valores são iguais.

Para ajustar o modelo de Brass a uma curva empírica ℓ_x , podem ser usados diferentes métodos. O primeiro passo do método tradicional é a conversão da série de ℓ_x numa série de $Y(x)$, usando a transformação logito para linearizar a relação conforme (20.12) e (20.13). Para estimar α e β , a abordagem mais óbvia seria fazer uma regressão de $Y(x)$ sobre $Y_p(x)$, mas Brass não recomenda esse procedimento porque ele tende a dar um peso excessivo às observações extremas. Em vez disso, propõe agrupar as séries $Y(x)$ e $Y_p(x)$ em dois conjuntos com números mais ou menos iguais de observações: um com os valores menores de $Y(x)$ e $Y_p(x)$ e outro com os valores maiores. Em cada um destes grupos se calcula a média dos valores observados, tanto de $Y(x)$ como de $Y_p(x)$. Assim se geram quatro médias $\bar{Y}_p, \bar{Y}_2, \bar{Y}_{p1}$ e \bar{Y}_{p2} . Agora α e β podem ser estimados da seguinte forma:

$$\beta = \frac{\bar{Y}_2 - \bar{Y}_1}{\bar{Y}_{p2} - \bar{Y}_{p1}} \quad \text{e} \quad \alpha = \bar{Y}_1 - \beta \bar{Y}_{p1} \quad (20.14)$$

O mesmo procedimento também é usado para o ajuste da fecundidade com a transformação de Gompertz (ver seção 20.4.2). Zaba e Paes (1995) propuseram um procedimento alternativo que opera diretamente sobre as TEMs e portanto pode ser aplicado também quando a mortalidade nos primeiros anos de vida é desconhecida. Mais em geral, tendo acesso aos recursos oferecidos por EXCEL, como o Solver, hoje em dia existem maneiras mais fáceis e mais precisas para estimar os parâmetros α e β do que (20.14). O método recomendado aqui começa com a passagem do conjunto de valores $Y_p(x)$ do Apêndice 3 para uma planilha em EXCEL. Na mesma planilha se incluem valores experimentais de α e β e uma tábua de vida preliminar baseada nos valores padrão da Apêndice 3 e os valores experimentais de α e β . Depois se usa o recurso Solver de EXCEL para encontrar os valores de α e β que minimizem a diferença entre os valores observados e ajustados de qualquer função desejada da tábua de vida.

Embora esses procedimentos possam ser usados para graduar tábuas de vida empíricas, na prática isso raramente é feito. Em países que dispõem de dados suficientemente detalhados para construir uma série empírica ℓ_x , o modelo de Brass geralmente é considerado demasiado grosseiro para fins de graduação e se usam outros modelos, como os discutidos nas seções anteriores. Onde

o modelo de Brass mostra a sua verdadeira utilidade é em lidar com situações onde a informação sobre mortalidade se limita a apenas uma indicação da mortalidade infantojuvenil e da mortalidade além dos 5 anos. Neste caso, esses dois dados podem ser convertidos em estimativas de α e β e depois, por meio da aplicação de (20.13), a uma tábua de vida estimada que reflete pelo menos as características principais do perfil etário da mortalidade no país.

Como exemplo, os seguintes dois valores caracterizam a tábua de vida (homens e mulheres) que foi derivada com base nos dados do Censo Demográfico de Moçambique de 2007:

$$\begin{aligned} \ell_5 &= 85.436 \\ \ell_{50} &= 57.374 \end{aligned} \quad (20.15)$$

Tomando como referência o padrão geral de Brass descrito no Apêndice 3, isso significa que

$$\beta = \frac{\logit(0,57374) - \logit(0,85436)}{Y_p(50) - Y_p(5)} = 1,2684 \quad (20.16.a)$$

$$\alpha = \logit(0,85436) - \beta Y_p(5) = -0,1217 \quad (20.16.b)$$

A partir destes valores de α e β é possível construir uma tábua de vida inteira. O procedimento COMBIN de MORTPAK também combina a mortalidade adulta com a infantojuvenil, mas se baseia nas tábuas de vida modelo de Princeton ou das Nações Unidas que serão discutidas na próxima seção. Este último procedimento foi usado para construir a tábua de vida oficial de Moçambique para o Censo de 2007.

Existem variações do modelo de Brass que tentam superar a sua rigidez, de modo que possam ser usados de uma forma mais apropriada para graduar dados de mortalidade baseados numa variedade mais ampla de padrões etários da mortalidade. Já se mencionou a possibilidade de usar padrões distintos $Y_p(x)$ para melhor cobrir certas variantes regionais específicas como a mortalidade africana. Houve tentativas para aumentar a flexibilidade do modelo, acrescentando parâmetros adicionais. A primeira proposta nesse sentido foi formulada por Zaba (1979), que acrescentou dois parâmetros. Mas o modelo resultante de quatro parâmetros é pouco usado na prática. Outro modelo de quatro parâmetros que foi proposto na época, mas que tampouco foi amplamente adotado, é o de Ewbank, Gómez de León e Stoto (1983).

O modelo mais recente de Murray et al. (2003) também possui quatro parâmetros (embora Moultrie et al., 2013 mostrem que efetivamente são apenas dois) e consiste em modificar o padrão geral de Brass por meio de duas funções de ajuste, chamadas γ_x e θ_x , que quantificam o desvio da mortalidade juvenil (até os 5 anos) e de adultos (sobrevivência até os 60 anos) em relação ao padrão geral. Isso dispensa o uso de padrões específicos como o padrão africano já que o ajuste por meio de γ_x cumpre o mesmo objetivo.

$$Y(x) = \logit(\ell_x) = \alpha + \beta Y_s(x) + \gamma(x) \left(1 - \frac{Y(5)}{Y_s(5)}\right) + \theta(x) \left(1 - \frac{Y(60)}{Y_s(60)}\right) \quad (20.17)$$

Esse modelo é usado atualmente pela Organização Mundial da Saúde (OMS) para graduar o seu conjunto de estimativas de tábuas de vida para todos os países do mundo (www.who.int/healthinfo/mortality_data/en/). Os detalhes desse modelo vão além do escopo deste livro, mas o Apêndice 4 reproduz os valores de γ_x , θ_x e $\ell^p(x)$. Os leitores interessados são referidos ao artigo original e a Moultrie et al. (2013), que assinalam uma inversão nos signos de γ_x e θ_x no artigo original, que é corrigida no Apêndice 4. É de notar-se que $\gamma_5 = \gamma_{60} = \theta_5 = \theta_{60} = 0$. Isso significa que é possível usar o mesmo procedimento para a estimação de α e β que foi descrito em (20.14), desde que as idades usadas para definir os dois pontos da curva sejam 5 e 60 anos. Os demais parâmetros são usados para otimizar o ajuste do resto da curva. Alternativamente, se pode usar o Solver de EXCEL ou outro tipo de otimizador para ajustar o modelo a qualquer função da tábua de vida. Outra possibilidade é o algoritmo ModMatch da OMS, em Stata (versão 7.0 ou mais), que pode ser descarregado da internet em <http://www.who.int/entity/healthinfo/statistics/modmatch.zip>.

20.2.3 Tábuas de vida modelo

A terceira abordagem para desenvolver modelos de mortalidade é pela síntese de tábuas de vida empiricamente observadas. A primeira tentativa neste sentido foram as tábuas de vida modelo (às vezes chamadas “tábuas-tipo”, por analogia ao francês, “tables-type”) publicadas pelas Nações Unidas na década de 50 (United Nations, 1955). Estas simplesmente consistiam nas médias das tábuas de vida observadas para um determinado nível de mortalidade, desagregadas por sexo. Portanto, estas tábuas de vida, que se baseavam predominantemente na experiência dos países mais desenvolvidos, só tinham um parâmetro, o nível de mortalidade. Sabendo a esperança de vida ou a mortalidade infantil de uma população se podia inferir toda a sua tábua de vida. Entretanto, não demorou para que ficasse evidente que esta conceptualização era demasiado simplista e que populações com a mesma esperança de vida podiam variar significativamente no seu padrão de mortalidade por idade. Ledermann e Bréas (1959) fizeram uma análise de componentes principais e identificaram pelo menos mais duas dimensões de variação, além do nível de mortalidade. Como se indicou na seção anterior, o mesmo problema de falta de flexibilidade surge no caso do modelo relacional onde dois parâmetros podem não ser suficientes. Entretanto, a solução adotada no caso das tábuas de vida modelo é diferente daquela adotada no caso do modelo relacional. Em vez de introduzir parâmetros adicionais para caracterizar dimensões específicas do padrão etário como a mortalidade infantil ou de crianças, a estratégia é identificar conglomerados ou “famílias” de tábuas de vida que exibem comportamentos parecidos em várias dimensões ao mesmo tempo, de forma semelhante com a adoção de padrões distintos (o geral, o africano etc.) no caso do modelo relacional.

O primeiro conjunto de modelos publicado com este intuito foram os modelos regionais de Princeton, que distinguem quatro famílias: “Oeste”, “Norte”, “Leste” e “Sul”. O conjunto original foi publicado nos anos 60 (Coale e Demeny, 1966) e houve uma atualização nos anos 80 (Coale e Demeny, 1983). As famílias se baseiam em 326 duplas (masculina e feminina) de tábuas de vida, maioritariamente da Europa e outros países desenvolvidos; não se usou nenhuma tábua africana. Apesar disso, elas continuam sendo aplicadas para fins de estimativas da mortalidade nos países em desenvolvimento. As tábuas do modelo “Norte” refletem principalmente a experiência dos países escandinavos, os do modelo “Sul” se baseiam nos países mediterrâneos e “Leste”, os países

do Leste europeu. O modelo “Oeste” é o mais geral e normalmente é usado quando não há razões para supor que o padrão de mortalidade tenha as características de um dos outros três. O modelo “Norte” se caracteriza por uma mortalidade infantil baixa e uma elevação da mortalidade nas idades entre 3 e 45 anos. O modelo “Leste” tem uma mortalidade infantil relativamente elevada que cai para níveis baixos nas idades subseqüentes, de 1 a 6 anos. O modelo “Sul” tem a mortalidade infantil mais elevada de todos, mas os níveis de mortalidade mais baixos entre os 7 e 70 anos. Depois dos 70 anos, o modelo “Norte” se caracteriza pela mortalidade mais baixa, enquanto “Leste” tem a mortalidade mais alta. Por exemplo, para a mesma esperança de vida ao nascer feminina de 65 anos, a mortalidade infantil varia entre 48,9 por mil no modelo “Norte” e 77,0 por mil no modelo “Sul”, enquanto a TEM aos 75 anos varia entre 58,8 por mil no modelo “Norte” e 72,0 no modelo “Leste”.

Tabela 20.1: Características principais da mortalidade e número de tábuas de vida empíricas usadas na construção dos quatro padrões regionais definidos por Coale e Demeny nas tábuas modelo de 1966

Características	Familia de Tabelas Modelo			
	Norte	Sul	Leste	Oeste
Infância	Baixa	Alta (menor de 5 anos)	Alta	*
Adultos	Alta	Baixa	Média	*
Idosos	Baixa	Alta (maior de 65 anos)	Alta	*
Tábuas de vida	9	23	31	130
Região de onde seriam as tabelas originais	Escandinávia	Mediterrâneo	Europa Central	Outros países desenvolvidos

Fonte: Paes (1996).

O desafio do uso dos modelos regionais é determinar qual é o modelo que melhor se adapta ao contexto do país sendo analisado. A aplicação mais frequente dos modelos de Princeton (e das Nações Unidas, a seguir) é a construção de tábuas de vida inteiras a partir do conhecimento da mortalidade infantil e juvenil. Como os métodos indiretos mais tradicionais (ver Capítulo 23) para a estimação da mortalidade se baseiam nas idades infantis e juvenis, muitas vezes a tábua de vida inteira é extrapolada a partir daí. Mas os resultados podem variar consideravelmente dependendo do modelo escolhido. Por exemplo, uma probabilidade de morte de 100 por mil entre o nascimento e a idade de 5 anos corresponde a uma esperança de vida masculina de 63,4 anos no modelo “Sul”, mas a apenas 58,9 anos no modelo “Oeste”.

Além da sua reedição em 1983, os modelos foram estendidos posteriormente. Coale e Guo (1989) estenderam as tábuas a níveis de mortalidade extremamente baixos, além da esperança de vida feminina de 77,5 anos que era o limite no sistema original. Por outro lado, a Divisão de População das Nações Unidas (<https://population.un.org/wpp/Download/Other/MLT/>) estendeu as tábuas a idades mais avançadas do que as consideradas na formulação original. Existem vários softwares que geram as tábuas de Princeton a partir da esperança de vida aos 10 anos (e_{10}), entre os quais o MORTPAK (ver seção 17.2) e demogR (ver seção 17.1 do Capítulo 17).

O sistema de tábuas de vida que foi desenvolvido pelas Nações Unidas nos anos 80 (Naciones Unidas, 1983) basicamente segue a mesma lógica dos modelos de Princeton, com a diferença de que se baseia mais em tábuas de vida de países em desenvolvimento. Em vez de quatro, há cinco

famílias de tábuas de vida, que se chamam “América Latina”, “Chile”, “Sul da Ásia”, “Extremo Oriente” e “Geral”. A diferenciação entre o modelo latino-americano e o chileno pode estranhar, mas basicamente se trata de uma diferença na mortalidade infantil e de crianças, sendo que o modelo chileno tem uma mortalidade infantil mais alta e uma mortalidade mais baixa daí até os 40 anos. O modelo “Extremo Oriente” possui a mortalidade mais baixa até os 13 anos, mas o nível mais alto a partir dos 15 anos. O modelo “Geral” cumpre a mesma função, de modelo “residual”, que o modelo “Oeste” tem no sistema de Princeton.

Tabela 20.2: A função ℓ_x para as tábuas de vida modelo de Princeton e Nações Unidas com $e_0 = 65$

Hom..	Oeste	Norte	Leste	Sul	Am. Lat.	Chile	Sul Ásia	Ext. Or.	Geral
0	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000
1	95,379	95,454	94,668	92,965	94,076	93,872	93,416	96,720	94,847
5	94,268	93,585	93,783	91,091	91,920	92,944	90,790	95,915	93,341
10	93,748	92,650	93,392	90,687	91,288	92,647	90,192	95,587	92,839
15	93,338	92,075	93,059	90,364	90,905	92,386	89,914	95,278	92,488
20	92,632	91,093	92,427	89,896	90,329	91,910	89,569	94,771	91,941
25	91,636	89,694	91,526	89,239	89,460	91,178	89,159	94,030	91,160
30	90,635	88,264	90,616	88,535	88,446	90,241	88,648	93,163	90,259
35	89,529	86,745	89,623	87,640	87,304	89,061	88,019	92,126	89,196
40	88,128	85,090	88,394	86,577	85,851	87,508	87,113	90,666	87,786
45	86,196	83,082	86,667	85,060	83,952	85,361	85,717	88,433	85,817
50	83,369	80,591	84,017	82,880	81,335	82,363	83,508	85,063	82,971
55	79,202	76,992	79,824	79,581	77,704	78,130	79,883	79,692	78,775
60	73,103	72,407	73,636	74,801	72,517	72,104	74,402	72,142	72,736
65	64,572	65,779	65,075	68,002	65,421	63,787	66,152	61,173	64,233
70	53,259	56,586	53,701	58,559	55,782	53,002	55,214	47,622	53,035
75	39,286	44,523	39,343	45,288	43,743	40,241	41,847	32,985	39,670
80	24,208	30,106	23,752	28,587	30,295	26,909	27,961	19,569	25,785
85	11,170	15,921	10,439	12,807	17,177	14,888	15,808	9,327	13,651
90	3,319	5,693	2,830	3,292	7,194	6,203	6,870	3,231	5,284
95	550	1,182	407	402	1,934	1,709	2,029	722	1,312
100	46	126	28	21	289	271	356	93	184

Mulh.	Oeste	Norte	Leste	Sul	Am. Lat.	Chile	Sul Ásia	Ext. Or.	Geral
0	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000
1	94,942	95,106	93,562	92,304	93,832	92,327	92,678	95,435	94,294
5	93,108	92,507	91,891	89,502	90,748	90,751	89,487	94,062	92,006
10	92,472	91,399	91,345	88,981	89,971	90,366	88,835	93,655	91,352
15	91,970	90,706	90,965	88,626	89,564	90,058	88,552	93,336	90,956
20	91,201	89,827	90,378	88,111	89,031	89,561	88,168	92,708	90,374
25	90,175	88,716	89,569	87,411	88,287	88,851	87,709	91,833	89,585
30	88,998	87,430	88,655	86,599	87,352	87,960	87,178	90,702	88,621
35	87,642	86,013	87,579	85,682	86,197	86,848	86,512	89,348	87,442
40	86,038	84,406	86,269	84,614	84,720	85,423	85,645	87,597	85,968
45	84,067	82,396	84,642	83,258	82,897	83,557	84,436	85,290	84,083
50	81,499	80,067	82,461	81,547	80,547	81,059	82,687	82,080	81,560
55	77,981	76,913	79,379	79,092	77,441	77,698	79,867	77,582	78,036
60	73,162	72,845	74,935	75,684	73,042	72,922	75,392	71,350	73,032
65	66,258	66,929	68,263	70,356	66,753	66,211	68,416	63,011	65,963
70	56,577	58,325	58,281	62,102	57,707	56,586	58,506	52,297	56,177
75	43,552	46,446	44,299	49,382	45,938	44,629	45,485	39,628	43,620
80	28,278	31,956	27,710	32,497	32,589	31,282	30,683	26,363	29,507
85	13,882	17,228	12,422	15,242	18,896	18,294	16,883	14,033	16,056
90	4,597	6,548	3,480	4,325	8,318	8,255	6,876	5,572	6,431
95	898	1,538	524	621	2,511	2,619	1,864	1,501	1,703
100	93	199	39	41	469	531	304	250	269

Fonte: Nações Unidas <https://population.un.org/wpp/Download/Other/MLT/>.

O programa MORTPAK, da Divisão de População das Nações Unidas, contém vários recursos para facilitar o manejo das tábuas de vida modelo de Princeton e das Nações Unidas. A conveniência principal é a disponibilidade das tábuas de vida modelo dentro do próprio software, evitando assim a busca da informação em tabelas publicadas. O procedimento BESTFT, por exemplo, ajuda o usuário a encontrar a tábua de vida modelo de Princeton ou das Nações Unidas que melhor se ajusta a uma determinada série de valores de ${}_n m_x$ ou ${}_n q_x$, enquanto CORMOR mostra os valores de uma determinada função da tábua de vida para todas as tábuas modelo que correspondem a um determinado parâmetro (por exemplo ℓ_{20}) especificado pelo usuário.

Os sistemas de tábuas de vida modelo de Princeton e das Nações Unidas descritos acima são os mais consolidados e vários dos métodos indiretos que serão descritos no Capítulo 23 foram desenvolvidos de forma integrada com esses modelos. Entretanto, existem outros sistemas de tábuas de vida como os da OCDE (Clairin et al., 1980) e o sistema que foi proposto por Frias e Rodrigues (1981) para modelar os padrões da mortalidade brasileira. Um modelo alternativo e mais atualizado para captar as especificidades do padrão da mortalidade no Brasil foi desenvolvido por

Campanário e Maia (2004). Paes (1993, 2014) desenvolveu um modelo para o Brasil, chamado “Modelo Adaptativo Brasileiro” e um sistema modelo específico para o Estado de São Paulo. Horta (2005) também investigou a existência de uma “norma” para o padrão da mortalidade brasileira e os eventuais desvios sistemáticos em relação a esta norma. Souza (2014) concluiu, a partir de um estudo dos padrões de mortalidade municipais no Brasil, que não existe um padrão único, mas em realidade seis padrões distintos para a mortalidade masculina e quatro para a feminina.

Mais recentemente foram desenvolvidos outros sistemas baseados em dados alternativos. O modelo de Wilmoth et al. (2012) tem dois parâmetros ($h = {}_5q_0$ e k) e se baseia nos dados da Human Mortality Database (HMD). A especificação do modelo é a seguinte:

$$\ln(m_x) = a_x + b_x \ln({}_5q_0) + c_x \ln({}_5q_0)^2 + v_x k \quad (20.18)$$

onde a_x , b_x , c_x e v_x são séries empíricas que derivadas dos dados da HMD. Elas são reproduzidas no Apêndice 5, separadamente para homens e mulheres. O parâmetro h ou ${}_5q_0$ expressa o nível de mortalidade e k modifica o padrão etário.

Outro sistema que precisa ser mencionado aqui são as tábuas de vida modelo desenvolvidas pela rede INDEPTH baseadas em informação dos Sítios/Áreas de Monitoramento Demográfico (ver Capítulo 5). Estas se destacam pelo fato de constituírem o único sistema de tábuas de vida baseado predominantemente em dados africanos e de tomar em conta o efeito da AIDS (SIDA) sobre os padrões de mortalidade por idade. Uma primeira versão do sistema foi publicado em 2004 (INDEPTH, 2004). Mais recentemente, o sistema foi revisado (Clark et al., 2012). O sistema de INDEPTH distingue entre quatro padrões. O Padrão 1 caracteriza os países africanos sem uma incidência significativa de AIDS (SIDA). O Padrão 2) descreve os países africanos onde a AIDS (SIDA) contribui significativamente à mortalidade. O Padrão 3) se aplica aos países asiáticos, com a exceção de Bangladesh. O Padrão 4) é o mais apropriado para Bangladesh, Indonésia e Vietnã. A desvantagem principal do sistema INDEPTH é que se baseia em relativamente poucas tábuas de vida de contextos altamente selecionados que podem não ser representativos para as suas regiões ou mesmo para os seus países.

20.3 MODELOS DE NUPCIALIDADE

Além do seu interesse intrínseco, as funções de nupcialidade servem como um dos elementos que definem alguns dos esquemas de fecundidade que serão introduzidos na próxima seção. O esquema mais conhecido para descrever a idade do primeiro casamento foi proposto por Coale e McNeil (1972), baseado na seguinte função padrão:

$$g_s(x) = 0,1946 \exp\{-0,174(x-6,06) - \exp[-0,2881(x-6,06)]\} \quad (20.19)$$

A função de Coale-McNeil agora consiste numa transformação linear simples de (20.19):

$$g(x) = \frac{C}{k} g_s\left(\frac{x - x_0}{k}\right) \quad (20.20)$$

onde C é a proporção de pessoas que em algum momento se casam (ou seja, 1 menos o celibato definitivo), x_0 é a idade inicial com que as pessoas começam a se casar e k é a inversa da rapidez com que o número de pessoas casadas aumenta.

Uma versão ligeiramente modificada de (20.20) foi proposta por Rodríguez e Trussell (1980):

$$g(x) = 1,2813 \frac{C}{k} \exp \left\{ -1,145 \left(\frac{x-m}{k} + 0,805 \right) - \exp \left(-1,896 \left(\frac{x-m}{k} + 0,805 \right) \right) \right\} \quad (20.21)$$

onde C e k têm o mesmo significado que têm em (20.20) e m se refere à idade média ao casar.

Além do problema de que os padrões de casamento estão em profunda transformação, uma crítica que pode ser feita a (20.20) e (20.21) é que esses padrões foram derivados em grande medida com base na experiência europeia e portanto não descrevem bem o comportamento das populações africanas ou latino-americanas. O terceiro modelo usado para descrever o padrão etário da nupcialidade é diferente na medida em que é totalmente paramétrico e não se baseia no ajuste a qualquer experiência empírica específica. O modelo de Hernes (1972) foi esquecido durante muitos anos mas ultimamente está recebendo uma atenção renovada devido ao fato de que se ajusta a uma variedade mais ampla de contextos. A forma mais transparente do modelo é em termos da seguinte equação diferencial:

$$\frac{d P_x}{d x} = A b^x P_x (1 - P_x) \quad (20.22)$$

onde o parâmetro A é parecido com C em (20.22) e descreve a intensidade da nupcialidade na população, b descreve a diminuição da propensão ao casamento com a idade e os dois fatores finais estabelecem uma relação entre a incidência de novos casamentos e a proporção P_x de pessoas que já estão casadas na idade x .

20.4 MODELOS DE FECUNDIDADE

20.4.1 Modelos puramente matemáticos

Um dos modelos mais simples que têm sido propostos para representar o perfil etário da fecundidade é o modelo de Brass, que possui apenas dois parâmetros: um (C) para o nível e um (s) para a localização da fecundidade por idade:

$$f(x) = C (x - s) (s + 33 - x)^2 \quad (20.23)$$

onde se entende que $f(x)=0$ para $x < s$ ou $x > s+33$. O parâmetro C indica o nível de fecundidade; mais especificamente $C = 12 TFT / 33^4$. Brass usou uma função de fecundidade desta forma para interpolar as razões P/F do seu método de ajuste da fecundidade (ver Capítulo 23). O mérito principal de (20.23) é a sua extrema simplicidade, que facilita os cálculos baseados nesta especificação. O número 33 se refere à amplitude típica do intervalo reprodutivo. Eventualmente, esse número pode ser substituído por um parâmetro, o que converte (20.23) num modelo de três, em vez de

dois parâmetros. Entretanto, mesmo com essa adaptação, (20.23) não se ajusta bem à variedade de padrões de fecundidade encontrados em situações reais.

Existem outros modelos de fecundidade baseados em expressões polinomiais. Por exemplo, Schmertmann (2003) usa splines quadráticos, dividindo as idades férteis em cinco intervalos:

1. Desde a idade inferior (α) até um ponto intermédio x_i entre α e a cúspide P , dado por $x_i = (1 - \min(0,75; 0,25 + 0,025(P - \alpha))) \alpha + \min(0,75; 0,25 + 0,025(P - \alpha)) P$;
2. Entre x_i e a cúspide P ;
3. De P até H , que é onde a curva atinge a metade da sua altura máxima R ;
4. De H até $(H + \beta)/2$, onde β é a idade fértil máxima;
5. Entre $(H + \beta)/2$ e β . Geralmente se supõe que $\beta = 50$, mas em algumas situações é preciso fazer pequenos ajustes. Supõe-se uma derivada igual a 0 nesse ponto final β .

Em cada um dos intervalos, as curvas consistem de polinômios de segundo grau que se juntam suavemente nos pontos de transição (nodos), com o mesmo valor e a mesma derivada. Assim, cada intervalo tem dois parâmetros livres que são estimados por meio de mínimos quadrados, a partir de dados anuais ou quinquenais de fecundidade observada. O autor mostra alguns procedimentos que facilitam esse cálculo. No total, o modelo possui 13 parâmetros. Devido a esse número relativamente grande de parâmetros, a sua utilidade principal é graduar taxas anuais ou desagregar taxas quinquenais. O seu uso em modelos de estimação da fecundidade do tipo P/F (ver Capítulo 23) é menos prático.

Outra alternativa é a interpolação com splines cúbicos que se baseia nos nodos fixos de 20, 25, 30 anos etc., exceto no caso do primeiro (α). Como a estrutura deste modelo é mais simples, com 7 em vez de 13 parâmetros, se presta mais a aplicações no contexto de métodos indiretos de estimação (ver Capítulo 23). Outra diferença é que se trata de um modelo de interpolação no sentido estrito, que reproduz os dados quinquenais, e não os ajusta como o modelo anterior. Um exemplo de como uma curva deste tipo pode ser estimada será mostrada na seção 23.3.2.4 do Capítulo 23.

Outro modelo que pode ser usado para descrever o padrão etário da fecundidade é a distribuição Gama, que possui muitas outras aplicações na estatística, e que é definida pela seguinte fórmula:

$$f(x) = TFT \frac{(x - d)^{\alpha-1} e^{-(x-d)/\beta}}{\beta^\alpha \Gamma(\alpha)} \text{ para } x > d \quad (20.24)$$

onde $\Gamma(\alpha)$ é a função gama definida como

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad (20.25)$$

A média da distribuição Gama é $\alpha\beta$ e a variância $\alpha\beta^2$. Tanto a função $\Gamma(\alpha)$ (20.25) como a própria distribuição (20.24), com $d=0$, fazem parte do acervo de funções disponíveis em EXCEL a partir da versão de 2013.

Finalmente, o padrão etário da fecundidade pode ser descrito por uma curva de Hadwiger, que tem a seguinte fórmula:

$$f(x) = \frac{C H}{S \sqrt{\pi}} \left(\frac{S}{x-d}\right)^{\frac{3}{2}} \exp\left(-H^2\left(\frac{S^2 + (x-d)^2}{S(x-d)} - 2\right)\right) \text{ para } x > d \quad (20.26)$$

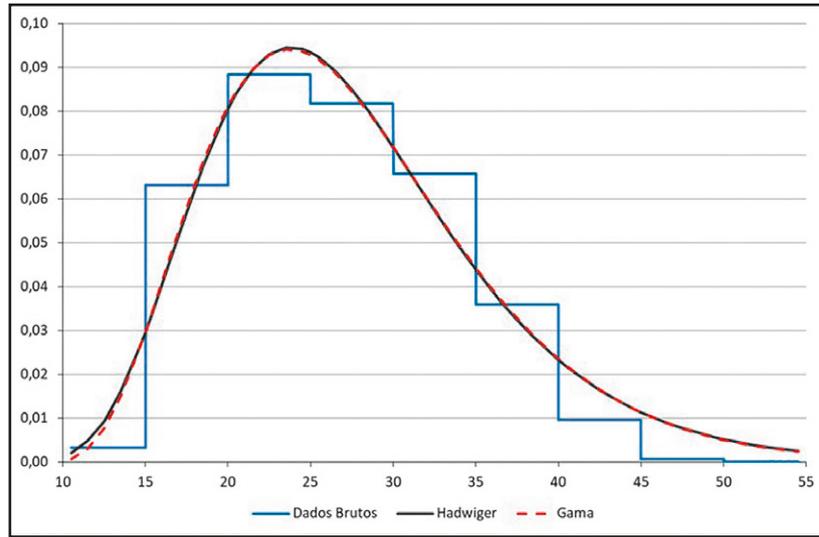
onde C se refere ao nível da fecundidade. Em realidade C é quase, embora não exatamente, igual à TFT. O parâmetro d determina a localização da curva e S a dispersão. O parâmetro H determina a inclinação da curva: para valores pequenos de H , a curva se inclina para a esquerda, enquanto valores maiores de H correspondem a uma curva mais simétrica. Uma versão simplificada da curva de Hadwiger usa apenas 3, em vez de 4 parâmetros, ao supor que $d = 0$.

O Gráfico 20.1 mostra os dados brutos da Tabela 10.2 e seu ajuste mediante as curvas Hadwiger e Gama. Os valores dos parâmetros usados nos ajustes são os seguintes:

Hadwiger:	C = 1,8212	Gama:	TFT = 1,8161	
	H = 2,2737		$\alpha = 4,9696$	$\Gamma(\alpha) = 22,9296$
	S = 27,52		$\beta = 3,7830$	
	d = 0		d = 8,7216	
				TFT = 1,8083

Estes valores foram determinados usando o critério de mínimos quadrados, supondo que a fecundidade de cada intervalo etário é representado pelo valor no seu ponto médio. Essa forma de graduação não é a ideal porque o suposto não é totalmente correto, mas a diferença é pequena e o procedimento de ajuste é bem mais fácil desta forma. Com a exceção do intervalo de 10-14 anos, as curvas são quase idênticas. O Gráfico 20.1 mostra uma das limitações principais de ambas as curvas, que em realidade se aplica a todas as curvas puramente matemáticas, ou seja que elas sobre-estimam a fecundidade no início e no final do intervalo etário. Mas no intervalo entre 15 e 40 anos o ajuste é bastante satisfatório.

Gráfico 20.1: Taxas Específicas de Fecundidade por grupo etário do Brasil, 2012-2014, brutas e ajustadas pelos modelos Hadwiger e Gama



Fonte: Dados brutos da Tabela 10.2.

Finalmente cabe mencionar outro modelo puramente matemático para representar funções de fecundidade, baseado na curva da distribuição normal. Na sua forma padrão, essa curva é unimodal (com um único pico) e simétrica. Evidentemente as curvas de fecundidade tipicamente não são simétricas, mas são unimodais. Ao representar o argumento da distribuição por um polinômio, é possível modificar a escala de idades, de tal forma que a curva resultante se ajuste ao padrão etário da fecundidade, ou seja,

$$f(x) = C \exp(-(A_n x^n + \dots + A_1 x + A_0)^2) \quad (20.27)$$

onde $A_n x^n + \dots + A_1 x + A_0$ é um polinômio *crescente* de grau n . No caso mais comum, onde as observações são dadas em intervalos de 5 anos, a escolha de $n=5$ geralmente permite fazer um ajuste exato (ou quase exato) da curva, ou seja, uma interpolação. Reduzindo o grau do polinômio a um n mais baixo, obtém-se uma curva que não representa os dados perfeitamente, ou seja, um ajuste no sentido estrito.

20.4.2 Coale-Trussell, Gompertz e Gompertz Relacional

As curvas puramente matemáticas introduzidas acima são usadas principalmente para a gradação de dados brutos de fecundidade. Embora algumas, como (20.23), tenham sido usadas para fins de modelação, as curvas mais frequentemente usadas em modelos de estimação do tipo que serão discutidos no Capítulo 23 se baseiam numa combinação de relações matemáticas com perfis etários que foram estabelecidos empiricamente.

Coale e Trussell (1974) formularam um modelo de fecundidade que decompõe a fecundidade em várias componentes: fecundidade natural, nupcialidade e controle da fecundidade marital.

$$f(x) = C G(x) n(x) e^{m v(x)} \quad (20.28)$$

O nível geral de fecundidade é representado pelo parâmetro C (ou M , como muitos autores preferem). A função $n(x)$ expressa os níveis observados da fecundidade natural encontrados por Henry (1961) nos seus estudos de 13 populações (principalmente históricas) supostamente sujeitas a um regime de fecundidade sem controles deliberados. As TEFs quinquenais² usadas no modelo de Coale-Trussell são as seguintes:

15-19 anos:	0,411
20-24 anos:	0,460
25-29 anos:	0,431
30-34 anos:	0,395
35-39 anos:	0,322
40-44 anos:	0,167
45-49 anos:	0,024

A função $v(x)$ representa a modificação da fecundidade natural devido ao controle da fecundidade, dependendo da paridade. O tamanho desta modificação é modulado pelo parâmetro m . Os valores quinquenais de $v(x)$, baseados em estudos empíricos de 43 populações, são os seguintes:

15-19 anos:	0
20-24 anos:	0
25-29 anos:	- 0,279
30-34 anos:	- 0,667
35-39 anos:	- 1,042
40-44 anos:	- 1,414
45-49 anos:	- 1,671

Finalmente, $G(x)$ é a versão cumulativa de (20.20), ou seja, a proporção de mulheres alguma vez casadas ou unidas até a idade x . Essa proporção também depende dos parâmetros x_0 e k definidos na seção 20.3. Como a forma matemática de (20.20) não permite a sua integração analítica, Coale e Trussell produziram tabelas para mostrar os resultados numéricos para diferentes valores de x , x_0 e k . Essas tabelas não são reproduzidas aqui; o leitor é referido ao artigo original (Coale e Trussell, 1974).

Ao longo do tempo foram formuladas várias críticas ao modelo de Coale e Trussell, particularmente ao uso de m como indicador do início do declínio da fecundidade marital. Também há um certo consenso de que os padrões etários de (20.28) não são muito adequados para modelar a

² O artigo original de Coale e Trussell também apresenta as taxas por idade simples, tanto de $n(x)$ como de $v(x)$.

fecundidade dos países da África Subsaariana. Xie e Pimentel (1992) criticaram o padrão etário da fecundidade marital e propuseram um padrão alternativo. Apesar dessas críticas, o modelo de Coale e Trussell continua sendo usado como referência em muitas técnicas de estimação indireta (ver Capítulo 23).

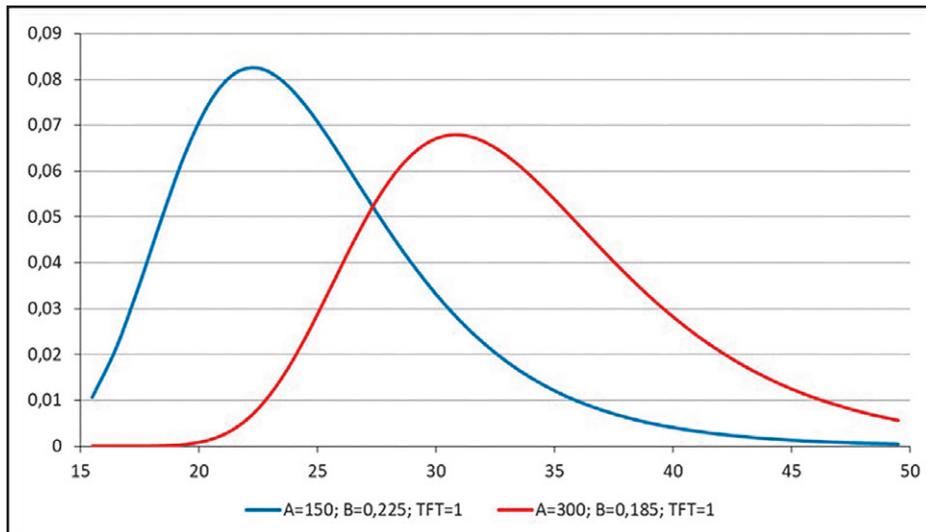
Com o objetivo de reduzir o número de parâmetros necessários para configurar a fecundidade por idade dos 3 propostos por Coale e Trussell, Brass (1981) propôs uma relação entre uma forma padrão de fecundidade e uma população qualquer.

A relação de Gompertz supõe que a proporção da fecundidade acumulada $F(x)$ segue uma distribuição com a seguinte forma:

$$F(x) = TFT \exp(-A \exp(-Bx)) \quad (20.29)$$

onde TFT é a Taxa de Fecundidade Total e A e B duas constantes. Diferenciando (20.29) obtém-se um perfil de fecundidade específica por idade do tipo mostrado no Gráfico 20.2. Como se pode apreciar no Gráfico 20.2, o parâmetro A move a localização horizontal da curva: quanto maior A , mais a curva se move para a direita. O parâmetro B determina a dispersão: valores maiores de B fazem a curva mais concentrada.

Gráfico 20.2: Perfil etário da fecundidade da curva paramétrica de Gompertz para duas combinações diferentes de parâmetros



Fonte: Elaboração própria.

A expressão (20.29) pode ser reduzida a uma forma linear a partir da seguinte transformação:

$$-\ln(-\ln(F(x) / TFT)) = -\ln(A) + Bx \quad (20.30)$$

Supondo que a função $F_p(x)$ representa a fecundidade acumulada segundo um esquema de fecundidade “padrão” e que $F_p(x)$ também se conforma à fórmula (20.30), tem-se que

$$-\ln(-\ln(F_p(x) / TFT_p)) = -\ln(A_p) + B_p x \quad (20.31)$$

Combinando (20.29) com (20.31) isso implica, para escolhas apropriadas de α e β , que

$$-\ln(-\ln(F(x) / TFT)) = \alpha + \beta (-\ln(-\ln(F_p(x) / TFT_p))) \quad (20.32)$$

A expressão da esquerda geralmente se nota como $Y(x)$ e a da direita como $\alpha + \beta Y_p(x)$, ou seja,

$$Y(x) = \alpha + \beta Y_p(x) \quad (20.33)$$

Para determinar a função $F(x)$ a partir de $Y(x)$ se usa a transformação inversa, ou seja

$$F(x) = TFT \exp(-\exp(-Y(x))) \quad (20.34)$$

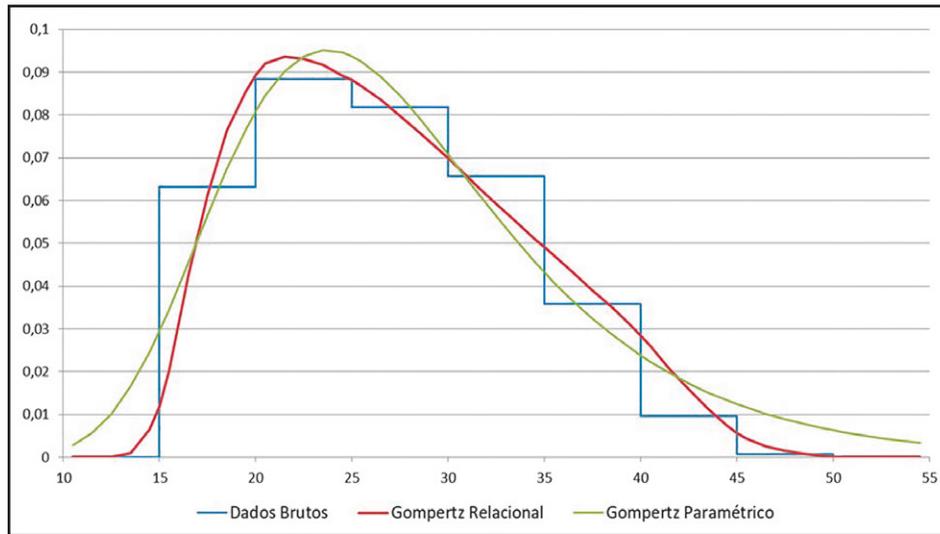
A vantagem de representar a curva de fecundidade no formato (20.32), em vez de (20.29) ou (20.30), é que (20.32) permite maior flexibilidade. Se $F_p(x)$ segue o padrão paramétrico de Gompertz expressado por (20.31), então (20.32) equivale a (20.29). Mas (20.32) também permite usar funções $F_p(x)$ que não se conformam ao padrão (20.30). Isso dá acesso a uma classe mais ampla de padrões etários que podem ajustar-se melhor à fecundidade observada do que o padrão paramétrico mais restritivo (20.29) ou (20.30). A fórmula (20.32) se chama a *fórmula relacional de Gompertz* e é bastante similar à fórmula de Brass (20.12) no caso da mortalidade, pois o objetivo de ambas é realizar uma transformação linear de uma distribuição para que ela possa ser comparada com uma distribuição padrão mediante o emprego de duas constantes α e β . De forma análoga ao *logito* no caso da fórmula de Brass (20.12), a função $-\ln(-\ln(x))$ em (20.30) é chamada de *gompito*.

Neste modelo, α é um indicador da localização do padrão de fecundidade em relação à idade enquanto que β é interpretado como o determinante da dispersão ou do grau de concentração do padrão de fecundidade. Em resumo pode-se dizer que, um valor negativo de α significa uma curva mais inclinada para a direita enquanto que um valor positivo indica uma curva mais voltada para a esquerda. Com relação a β , um valor maior que 1 é representado por uma curva mais estreita e um valor menor que 1, por uma curva mais larga.

Embora seja possível usar diferentes funções para o padrão $Y_p(x)$, na prática se recomenda o uso da função originalmente proposta por Booth (1984) e modificada por Zaba (Moultrie, 2013), a qual se reproduz no Apêndice 6. O padrão é adequado para populações de alta e média fecundidade, e com uma fecundidade acumulada normalizada (isto é, com TFT igual a um). Os valores do padrão $Y_p(x)$ são adquiridos como gompitos da função. O mesmo procedimento é feito para as parturições $Y_p(x)$, que são associadas a uma função de fecundidade padrão.

O Gráfico 20.3 compara as TEFs ajustadas do Brasil 2012-2014 (dados brutos da Tabela 10.2) usando o modelo de Gompertz relacional e paramétrico (20.26 e 20.29). O modelo relacional claramente fornece um ajuste mais próximo do que o paramétrico.

Gráfico 20.3: Taxas Específicas de Fecundidade por grupo etário do Brasil, 2012-2014, brutas e ajustadas pelos modelos Gompertz Relacional e Gompertz Paramétrico



Fonte: Dados brutos da Tabela 10.2.

Da mesma forma como no caso do modelo de mortalidade de Brass, existem diferentes métodos para ajustar uma curva empírica de fecundidade por idade ao modelo relacional de Gompertz. Tradicionalmente usa-se um método aproximado em que a fecundidade dos intervalos de 15-19, 20-24, 25-29 anos etc. é identificada com o ponto médio de cada intervalo. Um exemplo é mostrado na Tabela 20.3. A coluna $F(x)$ da mesma tabela multiplica as taxas da coluna anterior por 5 e acumula. O último número da mesma coluna é 2,504, o que corresponde à TFT. A próxima coluna faz a transformação gompito para linearizar $F(x)$. Depois se toma a média dos gompitos correspondentes aos intervalos de 15-19, 20-24 e 25-29 anos (0,540) e a média dos intervalos de 30-34, 35-39 e 40-44 anos (3,538). Essas médias são comparadas com as médias correspondentes da função de fecundidade padrão, na última coluna. Supondo que se usa o padrão do Apêndice 6 (dependendo da aplicação, podem ser usados outros padrões também), a primeira média pode ser obtida somando os valores de $Y_p(x)$ no Apêndice 6 para as idades de 15,5, 16,5, ..., 29,5 e dividindo por 15 (-0,378); a segunda média será para as idades de 30,5, 31,5, ..., 44,5 (1,657). Em seguida se estimam os parâmetros α e β que transformam os valores do padrão naqueles observados, de forma análoga a (20.14), ou seja:

$$\beta = \frac{3,538 - 0,540}{1,657 + 0,378} = 1,472 \quad e \quad \alpha = 0,540 - 1,472 (-0,378) = 1,097 \quad (20.35)$$

Tabela 20.3: Método tradicional de ajuste de uma curva de Gompertz relacional a dados empíricos de fecundidade

x	f(x)	F(x)	$-\ln(-\ln(F(x)/TFT))$	Média de três	Idem no Padrão
17,5	0,1094	0,547	-0,419		
22,5	0,1744	1,419	0,566	0,540	-0,378
27,5	0,1146	1,992	1,476		
32,5	0,0574	2,279	2,363		
37,5	0,0272	2,415	3,319	3,538	1,657
42,5	0,0142	2,486	4,931		
47,5	0,0036	2,504			

Encontrados os parâmetros α e β , se faz o caminho inverso, por meio de (20.33) e (20.34), para converter os gompitos ajustados em taxas de fecundidade convencionais, sempre usando 2,504 como a TFT. Por exemplo, a fecundidade acumulada para as idades de 15 e 16 anos será

$$F(15) = 2,504 \exp(-\exp(-(1,097 + 1,472 (-1,6225)))) = 0,066 \quad (20.36.a)$$

$$F(16) = 2,504 \exp(-\exp(-(1,097 + 1,472 (-1,3717)))) = 0,203 \quad (20.36.b)$$

etc.

As TEFs são obtidas por diferença entre a fecundidade acumulada de idades sucessivas; por exemplo, a TEF da idade simples de 16 anos seria $0,203 - 0,066$, ou seja, $0,137$.

Como no caso do modelo de Brass para a mortalidade, também existe um método de estimação mais direta, usando o Solver do EXCEL. Basta criar uma planilha de cálculo com os valores de $Y_p(x)$ do Apêndice 6 e definir três valores iniciais mais ou menos arbitrários (aqui escolhidos como 0, 1 e 2,5) para α , β e TFT. Na coluna C da Tabela 20.4 calculam-se os valores teóricos de $F(x)$ decorrentes destes valores de α , β e TFT, usando (20.33) e (20.34):

$$C2 = J\$3 * \text{EXP}(-\text{EXP}(-(\text{J\$1}+\text{J\$2}*\text{B2}))) \text{ e assim para baixo até a idade de 50 anos}$$

Na coluna F, a fecundidade acumulada da coluna C é convertida em TEFs para intervalos quinquenais:

$$F2 = (C12-C2) /5$$

$$F3 = (C22-C12)/5 \text{ e assim para baixo até F8}$$

Esses valores teóricos são comparados com os observados em G2:G8 e em G10 se calcula o desvio entre ambos:

$$G10 = (G2-F2)^2 + (G3-F3)^2 + (G4-F4)^2 + (G5-F5)^2 + (G6-F6)^2 + (G7-F7)^2 + (G8-F8)^2$$

Agora se chama o Solver de EXCEL, pedindo para minimizar G10 pelo ajuste de J1:J3. O resultado desta operação é $\alpha = 0,564$; $\beta = 1,399$ e $TFT = 2,463$.

Tabela 20.4: Método de ajuste de uma curva de Gompertz relacional a dados empíricos de fecundidade por meio do Solver de EXCEL

	A	B	C	D	E	F	G	H	I	J
1	Idade	$Y_p(x)$	$F(x)$			$5fx$	Observado		Alfa =	0
2	15	-1,7521	0,007825		15-19	0,066353	0,1094		Beta =	1
3	15,5	-1,6225	0,015773		20-24	0,12073	0,1744		TFT =	2,5
4	16	-1,4929	0,029197		25-20	0,115655	0,1146			
5	16,5	-1,3717	0,048521		30-34	0,093786	0,0574			
6	17	-1,2506	0,076067		35-39	0,067006	0,0272			
7	17,5	-1,1477	0,10703		40-44	0,030847	0,0142			
8	18	-1,0448	0,145651		45-49	0,004058	0,0036			
9	18,5	-0,9520	0,187384							
10	19	-0,8593	0,235695		Desvio		0,00792			
11	19,5	-0,7753	0,285089							
12	20	-0,6913	0,339589							
13	20,5	-0,6123	0,395194							
14	21	-0,5333	0,454636							
15	21,5	-0,4592	0,513494							
16	22	-0,3852	0,574866							
17	22,5	-0,3147	0,635364							
18	etc.									

Uma vez que o Solver tenha determinado os valores corretos dos parâmetros, as TEFs mostradas em F2:F8 já são as taxas ajustadas conforme o modelo relacional de Gompertz. A Tabela 20.4 não mostra as taxas ajustadas por idade simples, mas seria muito fácil incluir uma coluna adicional para já mostrá-las automaticamente. Note-se que as estimativas dos parâmetros obtidas acima são um pouco diferentes daquelas derivadas anteriormente, com o método tradicional, inclusive o parâmetro TFT. As estimativas obtidas com EXCEL devem ser consideradas mais precisas do que as anteriores, inclusive o parâmetro TFT. Entretanto, em casos onde se prefere usar a estimativa convencional de TFT, nada impede que se calcule J3 como 5 vezes a soma de G2:G8, aplicando o Solver apenas às células J1 e J2.

Vale mencionar ainda a existência de dois programas em “R”, de autoria de Eddie Hunsinger, que são mantidos na página web do Departamento De Demografia da Universidade de Berkeley (<https://applieddemogtoolbox.github.io/Toolbox/>), e que implementam o ajuste das curvas de mortalidade e fecundidade, usando o modelo logito e o modelo relacional de Gompertz.

20.5 MODELOS DE MIGRAÇÃO

O grau de formalização de esquemas modelo da migração é menor do que no caso da mortalidade e fecundidade. Em parte isso acontece porque a variabilidade dos padrões etários de migração é maior do que no caso da mortalidade e fecundidade. O modelo mais comum foi formulado por Rogers e Castro (1981) e consiste de quatro componentes que descrevem a migração por origem e destino (não a migração líquida). A componente principal está associada à migração nas idades economicamente ativas. A segunda componente descreve a migração nas idades iniciais, antes da entrada na força de trabalho. Em alguns países (mas nem sempre) é preciso incluir uma terceira componente, para descrever os movimentos associados à retirada da força de trabalho, por exemplo para as comunidade de aposentados no Sul dos EUA. Finalmente, o modelo inclui uma componente constante que se mantém ao longo da vida. As quatro componentes são expressadas pelas seguintes fórmulas:

$$\text{Idades Economicamente Ativas: } A_2 \exp\{-\alpha_2 (x-m_2) - \exp\{-\lambda_2 (x-m_2)\}\} \quad (20.37.a)$$

$$\text{Pré-entrada na Força de Trabalho: } A_1 \exp(-\alpha_1 x) \quad (20.37.b)$$

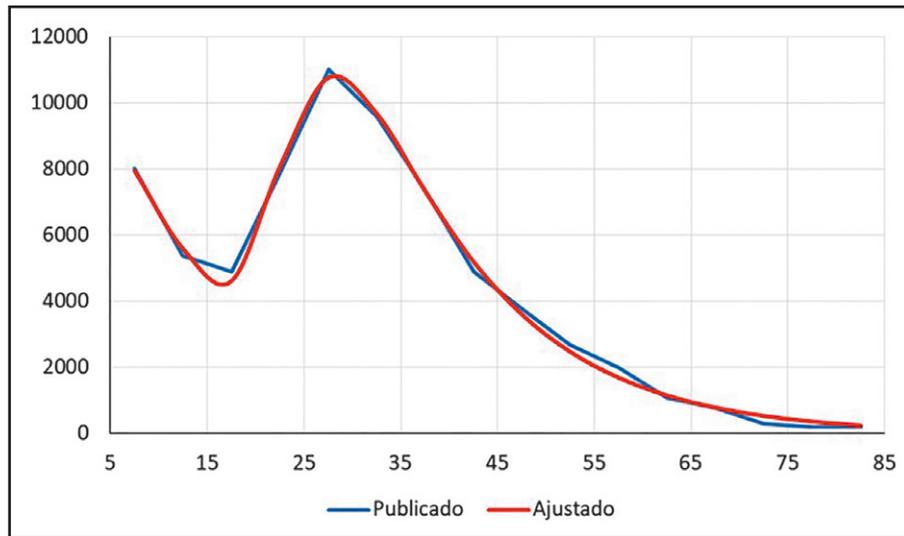
$$\text{Retirada da Força de Trabalho: } A_3 \exp\{-\alpha_3 (x-m_3) - \exp\{-\lambda_3 (x-m_3)\}\} \quad (20.37.c)$$

$$\text{Constante: } C \quad (20.37.d)$$

Como se pode notar, a primeira e terceira componente têm a mesma forma funcional, mas se localizam em pontos diferentes (m_2 e m_3) do ciclo vital. O número total de parâmetros necessários para ajustar o modelo é 4) para a primeira e terceira componente e 3) para as outras duas conjuntamente, ou seja, um total de 11. Nem todos os países possuem uma componente de migração associada à retirada da força de trabalho, tal como existe nos EUA e até certo ponto na Europa. Onde não há essa componente, o modelo pode ser simplificado, omitindo (20.37.c), o que reduz o número de parâmetros de 11 a 7.

Muitos padrões de fecundidade publicados foram graduados com o uso da função de Rogers e Castro, ainda que isso não seja explicitado. É o caso do Gráfico 11.1, como pode ser verificado ajustando uma curva de Rogers e Castro ao padrão etário da migração interprovincial masculina de 5 anos. Como se mostra abaixo no Gráfico 20.4, os dados coincidem quase perfeitamente com a curva ajustada.

Gráfico 20.4: Migração interprovincial de homens, Moçambique 2002-2007, segundo os dados publicados e graduados com a função de Rogers e Castro



Fonte: Gráfico 11.1 e cálculos próprios.

No Brasil, o modelo tem sido relativamente pouco usado. Beltrão e Henriques (1987) usaram os esquemas de Rogers e Castro, que foram desenvolvidos para analisar a migração unidirecional, em padrões etários aplicáveis à migração líquida. Jannuzzi (1998), com base na PNAD de 1993, elaborou uma série de padrões de migração e taxas migratórias-modelo para São Paulo, que permitiram aprofundar o estudo migratório no Estado. Machado (2005) estabeleceu curvas padrão para o Brasil, aplicando três métodos propostos por Rogers, Jordan e Raymer (2003). Santos (2019) elaborou padrões modelo por idade associando motivações migratórias e ciclos de vida.

