

TÉCNICAS DE INTERPOLAÇÃO

Embora a *suavização / graduação* de dados muitas vezes seja usada no mesmo contexto em que se aplica a *interpolação*, existe uma diferença fundamental entre as duas técnicas. Na suavização ou graduação de dados, procura-se corrigir os dados observados para obter tendências mais suaves que correspondem melhor a padrões teoricamente esperados. Portanto, a suavização *modifica* os dados originais. A interpolação, por outro lado, *mantém* os dados existentes, mas procura estimar os valores não observados de uma forma que seja consistente com os observados e que possui certas características desejáveis de suavidade e regularidade das tendências. Muitas vezes, ambas as técnicas são usadas conjuntamente. Por exemplo, no Capítulo 16, a informação sobre a população por idade simples foi agregada em intervalos quinquenais, para reduzir as variações erráticas frequentemente observadas nos dados completos. Um próximo passo poderia ser a redivisão dos dados quinquenais em idades simples por meio de uma interpolação. O resultado final seria uma *suavização* dos dados originais por meio de uma *agregação (agrupação)*, seguida por uma *interpolação*.

Uma outra distinção relevante é a diferença entre *interpolação* e *extrapolação*. Em, ambos os casos, os dados originais são mantidos. Mas na interpolação existem dados observados em ambos os lados do intervalo da interpolação, enquanto numa extrapolação existem dados observados apenas de um lado. Por exemplo, pode-se interpolar a população moçambicana entre os Censos de 2007 e 2017. Mas quando se procede além de 2017 (ou seja qual for o último censo disponível), entra-se no território das projeções, que no fundo são extrapolações com diferentes graus de sofisticação. Na prática, existe uma diferença grande entre interpolações e extrapolações devido às técnicas empregadas. As extrapolações geralmente se baseiam em fórmulas teóricas

para descrever os fenômenos estudados. Algumas dessas fórmulas serão discutidas no próximo capítulo. A interpolação, que forma o assunto deste capítulo, geralmente é feita com fórmulas matemáticas relativamente simples, de propósitos múltiplos, que não possuem uma justificativa teórica relacionada com o fenômeno sendo estudado. Por exemplo, as fórmulas de interpolação por meio de polinômios¹ são usadas tanto para interpolar dados de mortalidade como dados de fecundidade ou migração.

O objetivo da interpolação geralmente é a obtenção de alguma expressão analítica relativamente simples com as seguintes características:

1. A expressão precisa reproduzir os dados conhecidos. Por exemplo, se existem dois pontos de dados conhecidos (x_1, y_1) e (x_2, y_2) , o objetivo pode ser encontrar uma função $f(x)$ de tal forma que $f(x_1) = y_1$ e $f(x_2) = y_2$ e igualmente para os intervalos vizinhos.
2. Geralmente procura-se uma função relativamente regular e suave. No mínimo isso significa que $f(x)$ precisa ser contínua, mas de preferência também deve ser uma ou duas vezes diferenciável, principalmente nos pontos de junção entre um intervalo e outro, onde a especificação dos segmentos da função pode mudar.
3. Um dos desafios da interpolação tende a ser evitar que a função oscile muito. Como se verá abaixo, este é o problema fundamental de polinômios de graus elevados que funcionam bem em termos dos dois critérios anteriores, mas possuem uma tendência a oscilar excessivamente.
4. Frequentemente é preciso garantir que a função de interpolação possua outras propriedades. Por exemplo, muitas funções demográficas não permitem valores negativos e outras, por definição, precisam ser ascendentes ou descendentes.

A interpolação pode ser aplicada a

- Valores pontuais, como o tamanho da população em determinadas datas;
- Volumes, como a distribuição da população por faixa etária; ou
- Taxas / razões, tais como a TEM num determinado intervalo de idade.

Os procedimentos usados para cada finalidade são diferentes. O mais comum e mais simples é a interpolação de valores pontuais, por exemplo para estimar o tamanho da população do Brasil no dia 1º de julho de 1992, 1993, ..., 1999 com base na informação dos Censos de 1991 e 2000 e eventualmente de outros dados (os Censos de 1980 e 2010) também. Mais em geral, o problema pode ser formulado como a procura de uma função $f(x)$, com certos critérios de regularidade, de tal forma que $f(x_1) = y_1$, $f(x_2) = y_2$, ..., $f(x_n) = y_n$, onde x_1, x_2, \dots, x_n são os anos e y_1, y_2, \dots, y_n são valores conhecidos, como o tamanho da população segundo os respectivos censos. A grande maioria das fórmulas de interpolação foi desenvolvida para essa situação.

¹ Lembre-se de que um polinômio de grau n é uma função do tipo $A_n x^n + A_{n-1} x^{n-1} + \dots + A_1 x + A_0$.

Quando o objetivo for a interpolação de volumes associados a intervalos (também chamados “dados agrupados”), em vez de valores pontuais, ou de taxas, o procedimento é um pouco mais complexo. Um exemplo de volumes associados a intervalos é a interpolação da população com idades entre 5 e 10 anos exatos, para desagregá-la em intervalos de idades simples ou numa distribuição contínua da “densidade” de população entre 5 e 10 anos exatos.

Exemplos da interpolação de taxas incluem a desagregação de TEMs ou TEFs de intervalos quinquenais em taxas de idades simples ou em funções contínuas. Neste caso, o valor da taxa depende não só da intensidade do evento, mas até certo ponto também depende da distribuição da população subjacente. Por exemplo, a taxa de fecundidade adolescente pode ser significativamente mais alta se as mulheres de 15-19 anos se concentram mais no final do intervalo do que no caso oposto, de uma distribuição etária mais jovem. Embora muitas vezes sejam desconsiderados, em princípio esses efeitos precisam ser tomados em conta quando se faz a interpolação.

Embora não seja necessariamente o caso, a grande maioria das técnicas que serão discutidas neste capítulo especifica funções diferentes em segmentos / intervalos diferentes. O ponto onde muda a especificação de uma função para outra se chama um *nodo* ou *ponto nodal*. Na maioria das vezes os nodos correspondem aos pontos de interpolação ou aos pontos que separam os intervalos associados a dois volumes ou duas taxas distintas. Quase todas as técnicas apresentadas neste capítulo têm essa característica. Entretanto, existem técnicas, como a interpolação com B-splines na seção 18.2.3.3 ou o spline quadrático de Schmertmann na seção 18.2.3.5, que usam nodos em outros pontos que não coincidem com os pontos de interpolação ou com os limites entre intervalos de dados. Isso acontece mais frequentemente, embora não sempre, em casos onde o objetivo é graduação e não interpolação.

18.1 INTERPOLAÇÃO LINEAR

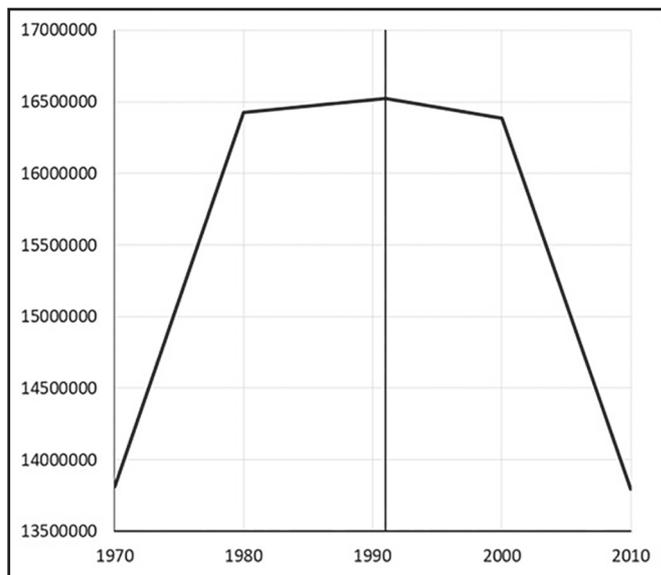
O tipo de interpolação mais simples, com o qual quase todos estão familiarizados, é a interpolação linear de valores pontuais $(x_1, y_1), \dots, (x_n, y_n)$, que é feita da seguinte forma:

$$f(x) = \frac{(x_{i+1} - x) y_i + (x - x_i) y_{i+1}}{x_{i+1} - x_i} \quad (18.1)$$

para o intervalo (x_i, x_{i+1}) e de forma análoga para os demais intervalos. Isso garante que $f(x_i) = y_i$ e $f(x_{i+1}) = y_{i+1}$. Graficamente (18.1) corresponde a conectar os pontos (x_i, y_i) com segmentos de linha retos. Uma característica importante da interpolação (18.1) é que o resultado para o intervalo (x_i, x_{i+1}) depende só de x_i , x_{i+1} , y_i e y_{i+1} e não dos valores de outros intervalos. Isso acontece porque (18.1) possui dois parâmetros (nível e inclinação do segmento de linha) e duas condições por satisfazer ($f(x_i) = y_i$ e $f(x_{i+1}) = y_{i+1}$). Portanto, não sobra nenhum parâmetro livre para garantir a regularidade da tendência entre intervalos sucessivos. Existem outros tipos de interpolação com essa característica (ver abaixo), mas na maioria dos casos existem parâmetros adicionais que podem ser escolhidos para otimizar a regularidade da função $f(x)$ entre intervalos. O Gráfico 18.1 mostra uma interpolação do tipo especificado em (18.1) com os dados da população de 0-4 anos nos Censos demográficos do Brasil de 1970, 1980, 1991, 2000 e 2010. Quando se usa EXCEL (com a opção “Dispersão com linhas retas”), o programa automaticamente exibe o

gráfico nesse formato linearmente interpolado, embora a linha de grade do ano 1991 tivesse que ser manualmente acrescentada.

Gráfico 18.1: Interpolação linear da população brasileira de 0-4 anos entre 1970, 1980, 1991, 2000 e 2010



Fonte: Censos Demográficos do Brasil.

A desvantagem da interpolação linear é que – devido à característica mencionada acima, de que cada intervalo é interpolado independentemente dos demais – o perfil resultante pode conter ângulos irrealistas como os observados em 1980 e 2000 quando a tendência da curva mudou. A evolução verdadeira da tendência seguramente foi bem mais suave. Para superar essa característica da interpolação linear existem alternativas como a interpolação com polinômios simples ou osculatórios ou splines, que serão discutidos mais adiante.

Mas antes de discutir essas alternativas, vale analisar como a interpolação linear pode ser aplicada no caso de volumes (dados agrupados) associados a intervalos. Uma forma frequentemente adotada, mas tecnicamente incorreta para fazer a interpolação de volumes é a identificação do ponto médio do intervalo de interpolação com a média do volume. Por exemplo, se o volume se refere à população de 15-19 anos, que tem um tamanho de 150.000 pessoas, o ponto médio do intervalo (17,5 anos) poderia ser identificado com o tamanho médio da população para cada idade, que é $150.000 / 5 = 30.000$ pessoas. Posteriormente estes pontos médios são interpolados linearmente, como em (18.1). O erro do procedimento consiste neste último passo, pois ao conectar os pontos médios com segmentos lineares se criam “quebradas” no meio dos intervalos e essas quebradas fazem com que o valor no ponto médio do intervalo não possa ser mais identificado com a população média para cada idade.

Uma solução possível e tecnicamente correta é acumular os valores associados aos intervalos, interpolar as somas parciais e diferenciar o resultado. Como será visto mais adiante, essa estratégia pode gerar resultados satisfatórios no caso de interpolações com curvas de graus mais elevados, mas no caso da interpolação linear o resultado acaba sendo absolutamente trivial: um volume

(geralmente uma população) homoganeamente distribuído ao longo de cada intervalo. Por outro lado, para que a própria densidade interpolada siga um perfil linear, o volume acumulado teria que ser interpolado com um polinômio do segundo grau. O seguinte procedimento oferece uma solução mais fácil para derivar densidades lineares dentro de cada intervalo, que são consistentes com os dados originais.

Identifica-se a população (ou outro tipo de volume) de cada intervalo com um valor pontual no meio do intervalo, a saber, a população média (o total dividido pela amplitude do intervalo), ou seja

$$f((x_i + x_{i+1})/2) = {}_n P_{xi} / n \quad (18.2)$$

onde $n = x_{i+1} - x_i$. Entretanto, em vez de conectar os pontos médios dos intervalos por meio de segmentos lineares (o que resulta em resultados inconsistentes com os dados originais), define-se um segmento linear dentro de cada intervalo, de tal forma que o valor no ponto médio é dado por (18.2) e as inclinações de cada segmento linear inicialmente ainda estão por determinar. Em realidade, só é preciso determinar a inclinação de *um* segmento linear, pois os demais seguem automaticamente. Por exemplo, supondo um valor inicialmente desconhecido para $f(x_1)$ e sabendo o valor $f((x_1+x_2)/2) = {}_n P_{x1} / (x_2 - x_1)$ no meio do intervalo, tem-se

$$f(x_2) = 2 {}_n P_{x1} / n - f(x_1) \quad (18.3)$$

onde $n = x_2 - x_1$. Mas sabendo $f(x_2)$ e $f((x_2+x_3)/2)$, sabe-se a inclinação do segundo segmento e também $f(x_3)$. Seguindo adiante da mesma forma, é possível construir a função f inteira. Portanto, a única incógnita verdadeira é o valor de $f(x_1)$. Qualquer escolha de $f(x_1)$ será consistente com a sequência de ${}_n P_{xi}$ observada, mas a escolha de $f(x_1)$ é determinada por algumas considerações:

Em alguns casos, há razões teóricas para escolher um determinado valor. Por exemplo, ao interpolar a fecundidade por idade, pode ser razoável supor que $f(14) = 0$ e derivar $f(15)$ com base nisso.

A escolha de $f(x_1)$ determina o grau de oscilação da sequência resultante, que precisa ser minimizado, otimizando a escolha de $f(x_1)$.

É preciso evitar que $f(x)$ assumam valores negativos. Particularmente, é preciso evitar valores negativos para $f(x_1)$ ou $f(x_n)$. Se a otimização de $f(x_1)$ resulta em valores negativos de um dos dois, é melhor reformular o problema, adotando valores alternativos x_1' e/ou x_n' para x_1 e x_n , de tal forma que $f(x) = 0$ para $x \leq x_1'$ e/ou $f(x) = 0$ para $x \geq x_n'$.

Ao utilizar o EXCEL, a determinação de $f(x_1)$, com base no segundo critério acima, é relativamente fácil de implementar. Os dados na Tabela 18.1 são do Censo do Timor-Leste de 2015 que contou 118.335 crianças nas idades de 1-4, 156.082 de 5-9, 156.269 de 10-14 e 136.872 de 15-19 (células B5:B8). O ponto inicial $f(x_1)$ consta na célula B2, inicialmente como um valor arbitrário. O mesmo valor deve ser reproduzido em D5. Os outros valores que aparecem na planilha foram gerados pelas seguintes fórmulas:

$$D2=(F5-F6)^2+(F6-F7)^2+(F7-F8)^2$$

$$\begin{array}{llll} C5=B5/4 & D5=B2 & E5=2*C5-D5 & F5=(E5-D5)/4 \\ C6=B6/5 & D6=E5 & E6=2*C6-D6 & F6=(E6-D6)/5 \\ C7=B7/5 & D7=E6 & E7=2*C7-D7 & F7=(E7-D7)/5 \\ C8=B8/5 & D8=E7 & E8=2*C8-D8 & F8=(E8-D8)/5 \end{array}$$

Tabela 18.1: Planilha de EXCEL para calcular a função de interpolação linear das populações de 1-4, 5-9, 10-14 e 15-19 anos do Censo do Timor-Leste, 2015

	A	B	C	D	E	F
1		Ponto Inicial		Oscilação		
2		28913,78		1231402		
3						
4	Idades	População	Ponto Médio	Ponto Inicial	Ponto Final	Inclinação
5	1-4	118335	29583,75	28913,78	30253,72	334,9842
6	5-9	156082	31216,4	30253,72	32179,08	385,0726
7	10-14	156269	31253,8	32179,08	30328,52	-370,113
8	15-19	136872	27374,4	30328,52	24420,28	-1181,65

Fonte: Censo Demográfico de Timor-Leste (2015).

O critério adotado aqui para minimizar as oscilações foi o valor inicial do primeiro intervalo que minimizou a soma das diferenças quadradas entre $F5$ e $F6$, $F6$ e $F7$, e entre $F7$ e $F8$, no caso um valor de 28913,78 para $f(x_p)$. A forma mais fácil de encontrar esse valor é por meio do recurso Solver do EXCEL, usando a célula D2 como objetivo e B2 como a célula variável.

A interpolação descrita acima, diferentemente daquela especificada em (18.1), implica certa interação entre os intervalos. Portanto, se a interpolação fosse realizada usando 5 ou 6 intervalos, em vez dos 4 usados acima, os resultados seriam diferentes (a não ser que o parâmetro A for escolhido de propósito para ser consistente com o resultado acima, em vez de minimizar as oscilações dos 5 ou 6 segmentos lineares).

A interpolação linear de taxas é um pouco mais complexa do que a interpolação de volumes e será tratada na seção 18.3.

Interpolação de formas funcionais linearizadas

A fórmula (18.1) pode ser generalizada para outras formas de interpolação, que não são lineares, mas que podem ser linearizadas mediante uma transformação. Aplica-se a transformação e em seguida a interpolação (18.1) e finalmente se aplica a transformação inversa. No Capítulo 9, por exemplo, foram mencionadas as fórmulas (9.7) e (9.8) que se baseiam na integração de curvas interpoladas com um perfil exponencial ou harmônico. Especificamente, o resultado (9.7) foi obtido da seguinte forma:

1. O intervalo de interpolação é interpretado como um intervalo que se estende de x a $x+n$, sendo que em x o valor da função é 1 e em $x+n$ o valor da função é ${}_n p_x$.
2. O ponto inicial e final são transformados com uma transformação logarítmica, para 0 e $\ln({}_n p_x)$, respectivamente.
3. Os valores transformados são interpolados linearmente.
4. A função resultante é transformada de volta com uma função exponencial. O resultado final é que $f(x+s) = {}_n p_x^{(s/n)}$.
5. Sobre esta expressão se aplica a seguinte operação de integração:

$${}_n a_x = \int_0^n (s-x) {}_n p_x^{s/n} ds / \int_0^n {}_n p_x^{s/n} ds \quad (18.4)$$

O resultado é o mostrado em (9.7).

A interpolação com outros perfis funcionais funciona da mesma forma, só que as transformações são diferentes. Por exemplo, para reproduzir a fórmula (9.6) o passo 2) usa uma função inversa ($1/x$) em vez de uma transformação logarítmica e o passo 4) também usa uma função inversa, esta vez para substituir a função exponencial usada acima. Transformações deste tipo são comuns em várias interpolações e extrapolações aplicadas na demografia. As fórmulas (21.1) a (21.5) ilustram esse procedimento no caso da extrapolação de tendências que seguem um perfil logístico ou baseado na função de Gompertz.

18.2 INTERPOLAÇÕES QUE PASSAM SUAVEMENTE DE UM INTERVALO PARA OUTRO

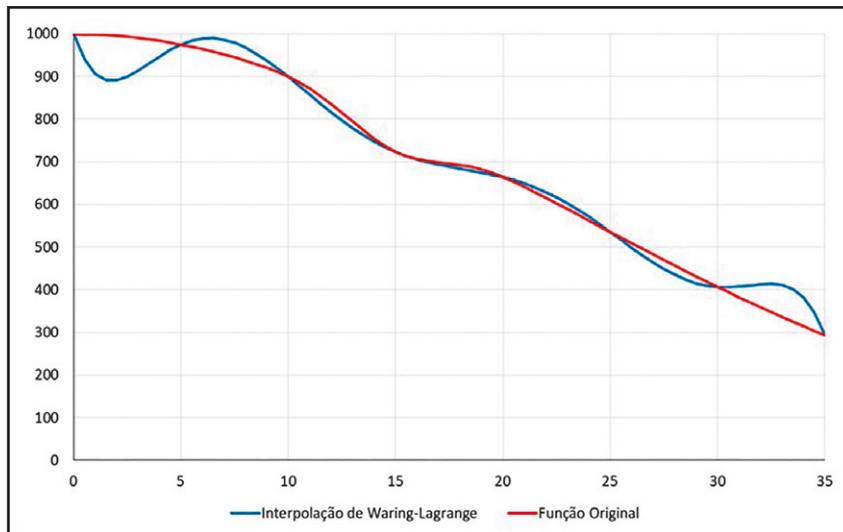
À primeira vista, o procedimento mais indicado para interpolar mais de dois valores pontuais de forma suave parece ser o uso de um polinômio de grau $n-1$, onde n simboliza o número de pontos. É sempre possível encontrar um polinômio desse tipo que passa por todos os pontos e o resultado é n vezes diferenciável antes que a derivada se reduza a 0. Uma expressão que define um polinômio de grau $n-1$ que passa pelos pontos $(x_1, y_1), \dots, (x_n, y_n)$ é a fórmula de Waring-Lagrange, que tem a seguinte forma:

$$f(x) = y_1 \frac{(x-x_2)(x-x_3)\dots(x-x_n)}{(x_1-x_2)(x_1-x_3)\dots(x_1-x_n)} + y_2 \frac{(x-x_1)(x-x_3)\dots(x-x_n)}{(x_2-x_1)(x_2-x_3)\dots(x_2-x_n)} + \dots + y_n \frac{(x-x_1)(x-x_2)\dots(x-x_{n-1})}{(x_n-x_1)(x_n-x_3)\dots(x_n-x_{n-1})} \quad (18.5)$$

Trata-se de uma fórmula que antigamente se aplicava relativamente pouco, pela dificuldade do cálculo, mas hoje em dia não é difícil programá-la em EXCEL. Entretanto, as curvas obtidas por meio das interpolações baseadas em (18.5) – quando são de graus elevados – têm a tendência a exibir oscilações indesejáveis, uma tendência conhecida como o *Fenômeno de Runge*. O problema é ilustrado no Gráfico 18.2 que mostra uma curva suave e uniformemente descendente que nos

pontos $x = 0, 5, 10, 15, \dots, 35$ é interpolada pelo polinômio de Waring-Lagrange, do 7º grau. O polinômio efetivamente reproduz os valores corretos nesses pontos, mas o seu comportamento para valores de x inferiores a 10 ou superiores a 30 se distancia significativamente da curva original, com oscilações implausíveis e inclusive inconsistentes com a característica descendente da curva original. Isso ilustra que o uso de polinômios de graus elevados para representar a curva original pode levar a erros significativos.

Gráfico 18.2: Ilustração do Fenômeno de Runge com uma função suave e descendente, interpolada nos pontos $x=0, 5, 10, 15, \dots, 35$ pelo polinômio de Waring-Lagrange



Fonte: Elaboração própria.

Pelo motivo explicado acima geralmente se evita o uso de polinômios de graus elevados. Em vez disso se usam polinômios de graus mais baixos – ou eventualmente outros tipos de funções – que se juntam de forma suave nos pontos de junções entre um intervalo e outro. As duas técnicas mais conhecidas para esse propósito são a interpolação *osculatória* e a interpolação com *splines*. As duas técnicas são semelhantes, mas a interpolação com splines é um pouco mais moderna do ponto de vista do uso de técnicas matemáticas.

Todos os exemplos de funções de interpolação osculatórias e splines discutidos aqui compartilham uma característica importante e conveniente com (18.5) que é a sua *linearidade*, não no sentido de que as próprias funções sejam lineares, mas no sentido de que podem ser combinadas de forma linear. Ou seja, se a função $f(x)$ interpola os pontos $(x_1, y_1), \dots, (x_n, y_n)$ e a função $g(x)$ interpola os pontos $(x_1, z_1), \dots, (x_n, z_n)$, os pontos $(x_1, A y_1 + B z_1), \dots, (x_n, A y_n + B z_n)$ podem ser interpolados pela combinação $A f(x) + B g(x)$. Uma consequência importante dessa propriedade é que basta derivar as funções de interpolação para os seguintes pontos para poder interpolar qualquer sequência de n valores nos pontos x_1, \dots, x_n pela simples multiplicação e soma das funções que interpolam cada uma das sequências:

$$y_1 = 1, y_2 = 0, y_3 = 0, \dots, y_n = 0 \quad \text{interpolados por } f_1(x) \quad (18.6.a)$$

$$y_1 = 0, y_2 = 1, y_3 = 0, \dots, y_n = 0 \quad \text{interpolados por } f_2(x) \quad (18.6.b)$$

$$\dots\dots\dots$$

$$y_1 = 0, y_2 = 0, y_3 = 0, \dots, y_n = 1 \quad \text{interpolados por } f_n(x) \quad (18.6.n)$$

Por exemplo, na fórmula (18.5) seria possível definir

$$f_1(x) = \frac{(x - x_2)(x - x_3) \dots (x - x_n)}{(x_1 - x_2)(x_1 - x_3) \dots (x_1 - x_n)} \quad (18.7.a)$$

$$f_2(x) = \frac{(x - x_1)(x - x_3) \dots (x - x_n)}{(x_2 - x_1)(x_2 - x_3) \dots (x_2 - x_n)} \quad (18.7.b)$$

e assim por diante, de modo que $f(x) = y_1 f_1(x) + y_2 f_2(x) + y_3 f_3(x) + \dots + y_n f_n(x)$. As funções $f_1(x), \dots, f_n(x)$ são chamadas as *funções de base*, no caso para a interpolação de Waring-Lagrange. O mecanismo funciona igual no caso da interpolação com polinômios osculatórios ou splines, só que as funções de base $f_1(x), \dots, f_n(x)$ são diferentes.

Antes de proceder com a discussão de fórmulas específicas de interpolação, vale discutir o que exatamente se entende por “suavidade” e “ausência de oscilações”. Estes critérios são satisfeitos de formas diferentes em diferentes métodos de interpolação. Os critérios mais comuns são os seguintes:

A minimização da expressão:

$$I_1 = \int_{x_1}^{x_n} (f''(x))^2 dx \quad (18.8)$$

Se $f(x)$ descreve uma função linear, (18,8) é igual a zero. Para polinômios de graus mais elevados, (18.8) tende a aumentar com o grau. A expressão (18.8) é usada, por exemplo, para caracterizar a otimalidade do spline natural (ver mais adiante). Entretanto, (18.8) não caracteriza perfeitamente o grau de curvatura de uma função. Em trechos da curva onde $f(x)$ aumenta ou diminui rapidamente, (18.8) exagera o grau de curvatura. Uma expressão que descreve a curvatura mais corretamente é a seguinte:

$$I_2 = \int_{x_1}^{x_n} (f''(x))^2 / (1 + f'(x)^2)^{2,5} dx \quad (18.9)$$

O problema desta expressão é que ela é difícil de aplicar em aplicações concretas, razão pela qual ela é pouco usada. Šalkauskas (1984) propõe um procedimento de splines naturais ponderados com base em (18.9), para melhorar o desempenho dos splines naturais comuns, mas este procedimento não será discutido neste capítulo.

Em vez de usar o critério (18.8), que exige que a função seja duas vezes diferenciável, às vezes é melhor expressar a regularidade como o comprimento total do arco da linha interpolada,

já que uma linha muito comprida indica oscilações desnecessárias. A fórmula para o comprimento do arco é

$$I_3 = \int_{x_1}^{x_n} \sqrt{1 + (f'(x))^2} dx \quad (18.10)$$

Infelizmente esta expressão geralmente é difícil de calcular analiticamente. Por isso, merece a preferência calcular o comprimento aproximado de forma numérica:

$$I_3 = \sum \sqrt{\Delta x^2 + (f(x + \Delta x) - f(x))^2} \quad (18.11)$$

onde x tem incrementos sucessivos de Δx . O critério de comprimento total do arco deve ser usado cuidadosamente porque nem sempre funciona bem. Por exemplo, tende a fornecer resultados melhores na interpolação valores pontuais do que na interpolação de volumes ou taxas, e pode não dar bons resultados para valores insuficientemente precisos de Δx .

Um critério final, que em realidade é um dos mais antigos, é a minimização das diferenças finitas de ordem k . A *primeira* diferença de uma função f entre x_1 e x_2 é definida como

$$\Delta^1(f, x_1, x_2) = (f(x_2) - f(x_1)) / (x_2 - x_1) \quad (18.12.a)$$

De forma análoga, a *segunda* diferença de f entre x_1 , x_2 e x_3 é definida como

$$\Delta^2(f, x_1, x_2, x_3) = 2 (\Delta^1(f, x_2, x_3) - \Delta^1(f, x_1, x_2)) / (x_3 - x_1) \quad (18.12.b)$$

e a *terceira* diferença como

$$\Delta^3(f, x_1, x_2, x_3, x_4) = 3 (\Delta^2(f, x_2, x_3, x_4) - \Delta^2(f, x_1, x_2, x_3)) / (x_4 - x_1) \quad (18.12.c)$$

É possível continuar desta forma para diferenças de ordens mais elevadas. Na grande maioria dos casos, estas fórmulas são aplicadas para pontos x_1 , x_2 , x_3 etc. separados por uma unidade, o que as simplifica um pouco, dispensando-se dos denominadores e dos coeficientes iniciais. Ao tomar todas as diferenças de uma determinada ordem k , elevando-as ao quadrado e somando os resultados, obtém-se outra medida de regularidade, a *soma dos quadrados das diferenças* de ordem k . Esta medida é relevante no contexto do método de Beers (ver abaixo), que minimiza a soma dos quadrados das quintas diferenças.

18.2.1 Curvas osculatórias: o método de Sprague

O termo “osculatório” vem do latim *osculum* (beijo), para simbolizar que as funções usadas para interpolar intervalos sucessivos têm transições suaves (como se se beijassem). Existem várias fórmulas desse tipo, mas aqui será apresentada apenas a interpolação de Sprague, que usa polinômios de 5º grau cujos coeficientes vão variando entre intervalos sucessivos. Supondo que há 6 valores para interpolar, $(x_1, y_1), \dots, (x_6, y_6)$, o polinômio do intervalo central, de (x_3, y_3) a (x_4, y_4) , é interpolado da seguinte forma:

1. Primeiro se define um polinômio de 4º grau, da esquerda, para interpolar os pontos (x_1, y_1) a (x_3, y_3) . Noutras palavras, se define um polinômio $f_{esq.}$ de quarto grau, com a propriedade de que $f_{esq.}(x_1) = y_1, \dots, f_{esq.}(x_3) = y_3$. Há várias maneiras para determinar os coeficientes de um polinômio de grau n que passa por $n+1$ pontos. A fórmula de Waring-Lagrange (18.5) é uma possibilidade. Mas em EXCEL ou em “R” pode ser mais fácil explicitar as $n+1$ equações que determinam os y_i como funções de x_1, \dots, x_{n+1} e resolver este sistema de equações para obter os coeficientes.
2. Em seguida, se define um polinômio de 4º grau, da direita, para interpolar os pontos (x_2, y_2) a (x_6, y_6) . Noutras palavras, se define um polinômio $f_{dir.}$ de quarto grau, com a propriedade de que $f_{dir.}(x_2) = y_2, \dots, f_{dir.}(x_6) = y_6$.
3. Para ambos os polinômios se calculam as derivadas de primeira e segunda ordem em x_3 e x_4 .
4. Agora o polinômio f de 5º grau para interpolar entre x_3 e x_4 é determinado pelas condições de ser igual a y_3 em x_3 , y_4 em x_4 , e de ter as derivadas do polinômio da esquerda em x_3 e as derivadas do polinômio da direita em x_4 . Noutras palavras: $f(x_3) = y_3, f'(x_3) = f_{esq.}'(x_3), f''(x_3) = f_{esq.}''(x_3), f(x_4) = y_4, f'(x_4) = f_{dir.}'(x_4)$ e $f''(x_4) = f_{dir.}''(x_4)$.
5. Agora, supondo que há mais de 6 pontos, o mesmo procedimento pode ser aplicado para $(x_2, y_2), \dots, (x_7, y_7)$, com o intervalo central de x_4 a x_5 . Os polinômios em ambos os intervalos serão diferentes, mas os seus valores e primeiras e segundas derivadas em x_4 serão os mesmos.
6. Se existem mais de seis pontos para interpolar, se procede da mesma forma nos demais intervalos, com a exceção dos primeiros dois e dos últimos dois, onde são usados os polinômios de 4º grau que serviram para a interpolação dos intervalos de x_3 a x_4 e de x_{n-3} a x_{n-2} , respectivamente.

Não é necessário que todos os intervalos sejam do mesmo tamanho, mas no caso mais comum onde todos têm a mesma amplitude a e com x_3 definido como o ponto 0, ou seja se $x_1 = -2a, x_2 = -a, \dots, x_6 = 3a$, as funções de base no intervalo entre x_3 e x_4 podem ser convenientemente escritas como²

² Siegel e Swanson (2004) apresentam as fórmulas diferentemente. A diferença reside no fato de que (18.8.a-f) são mais convenientes para a programação de rotinas em EXCEL. As tabelas apresentadas em Siegel e Swanson eram mais convenientes na época em que não se tinha um acesso fácil a computadores e o mais fácil era apresentar os valores mais relevantes em tabelas padronizadas ou com o uso de diferenças finitas.

$$f_1(x) = (-5 (x/a)^5 + 13 (x/a)^4 - 9 (x/a)^3 - (x/a)^2 + 2 (x/a)) / 24 \quad (18.13.a)$$

$$f_2(x) = (25 (x/a)^5 - 64 (x/a)^4 + 39 (x/a)^3 + 16 (x/a)^2 - 16 (x/a)) / 24 \quad (18.13.b)$$

$$f_3(x) = (-50 (x/a)^5 + 126 (x/a)^4 - 70 (x/a)^3 - 30 (x/a)^2 + 24) / 24 \quad (18.13.c)$$

$$f_4(x) = (50 (x/a)^5 - 124 (x/a)^4 + 66 (x/a)^3 + 16 (x/a)^2 + 16 (x/a)) / 24 \quad (18.13.d)$$

$$f_5(x) = (-25 (x/a)^5 + 61 (x/a)^4 - 33 (x/a)^3 - (x/a)^2 - 2 (x/a)) / 24 \quad (18.13.e)$$

$$f_6(x) = (5 (x/a)^5 - 12 (x/a)^4 + 7 (x/a)^3) / 24 \quad (18.13.f)$$

Estas são as fórmulas para a interpolação de valores pontuais. Para desagregar valores associados a intervalos em densidades, (18.13.a-f) precisam ser diferenciados. Por exemplo, para derivar uma função de densidade que caracteriza um volume igual a 1 no intervalo $(0,1)$ e 0 em todos os outros intervalos, é preciso diferenciar uma densidade acumulada igual a 0 em $x=-2$, 0 em $x=-1$, 0 em $x=0$, 1 em $x=1$, 1 em $x=2$ e 1 em $x=3$. Ou seja, há que diferenciar a soma $f_4(x) + f_5(x) + f_6(x)$. Como $a=1$, o resultado é

$$\frac{d}{dx} f_4(x) + f_5(x) + f_6(x) = (150 x^4 - 300 x^3 + 120 x^2 + 30 x + 14) / 24 \quad (18.14)$$

É preciso notar duas coisas em relação a (18.14):

A função de densidade acumulada é 0 em $x=0$ e 1 em $x=1$, mas a densidade (18.14) tem um valor de $14/24 = 7/12 = 0,58333\dots$ em $x=0$ e $x=1$.

A função de densidade acumulada é *duas* vezes diferenciável nos pontos de transição $x=0$ e $x=1$, mas (18.14) só é *uma* vez diferenciável nesses pontos.

Se os intervalos forem de tamanhos diferentes, as expressões (18.13.a-f) precisam ser modificadas. O seguinte programa em “R” gera os coeficientes de f_1, \dots, f_6 para intervalos de tamanhos arbitrários: x é o vetor (x_1, x_2, \dots, x_6) e y é o vetor de valores para interpolar (y_1, y_2, \dots, y_6) . Para obter os coeficientes de $f_1(x)$, escolha $y = (1,0,0,0,0,0)$, para obter os coeficientes de $f_2(x)$, escolha $y = (0,1,0,0,0,0)$ e assim adiante.

```
sprague <- function(x,y) {
  # Função para calcular uma interpolação de Sprague com 5 intervalos de tamanhos
  # arbitrários
  col1 <- x^4; col2 <- x^3; col3 <- x^2; col4 <- x; col5 <- x^0
  mat <- cbind(col1,col2,col3,col4,col5)
```

```

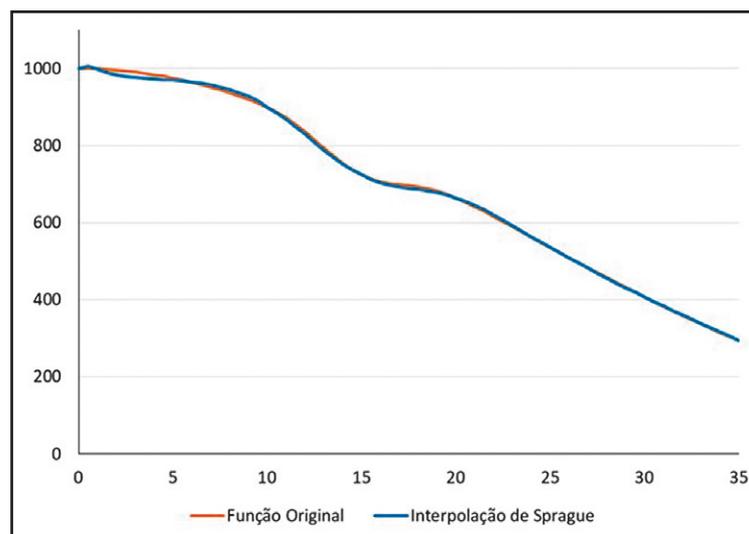
mat1 <- mat [1:5,1:5]; vec1 <- y [1:5]; dim(vec1) <- c(5,1)
mat2 <- mat [2:6,1:5]; vec2 <- y [2:6]; dim(vec2) <- c(5,1)
res1 <- solve(mat1,vec1); res2 <- solve(mat2,vec2)
mat3 <- array(c(x[3]^5,5*x[3]^4,20*x[3]^3,x[4]^5,5*x[4]^4,20*x[4]^3,
  x[3]^4,4*x[3]^3,12*x[3]^2,x[4]^4,4*x[4]^3,12*x[4]^2,x[3]^3,3*x[3]^2,
  6*x[3],x[4]^3,3*x[4]^2,6*x[4],x[3]^2,2*x[3],2,x[4]^2,2*x[4],2,x[3],
  1,0,x[4],1,0,1,0,0,1,0,0), dim=c(6,6))
vec3 <- array(c(y[3],4*res1[1]*x[3]^3+3*res1[2]*x[3]^2+2*res1[3]*x[3]
  +res1[4],12*res1[1]*x[3]^2+6*res1[2]*x[3]+2*res1[3],y[4],
  4*res2[1]*x[4]^3+3*res2[2]*x[4]^2+2*res2[3]*x[4]+res2[4],
  12*res2[1]*x[4]^2+6*res2[2]*x[4]+2*res2[3]),dim=c(6,1))
res3 <- solve(mat3,vec3)
res3
}

```

(18.15)

O uso de polinômios de graus mais baixos (4° ou 5°) reduz a sensibilidade dos resultados a oscilações. Como se mostra no Gráfico 18.3, uma interpolação com a curva de Sprague consegue melhorar dramaticamente a correspondência entre a curva original e os valores interpolados em comparação com o Gráfico 18.2.

Gráfico 18.3: A mesma função suave e descendente do Gráfico 18.2, interpolada nos pontos $x=0, 5, 10, 15, \dots, 35$ pela fórmula de Sprague



Fonte: Elaboração própria.

18.2.2 O método de Beers

O método de Beers tem certas semelhanças com o método de Sprague, mas a sua racionalidade é bastante diferente. A semelhança consiste na definição dos polinômios de 4º grau $f_{esq.}$ e $f_{dir.}$ em cada intervalo, que é feita da mesma forma como no método de Sprague. Entretanto, o critério de otimalidade é diferente. O método de Sprague se foca na transição suave das curvas que caracterizam diferentes segmentos, com diferenciabilidade até o segundo grau nos nodos. O método de Beers (1944) exige apenas que os valores interpolados nos nodos sejam os mesmos para os valores interpolados dos segmentos vizinhos, mas foca na regularidade da função f nos outros pontos interpolados. Tipicamente o método é usado para dividir intervalos quinquenais em intervalos anuais. Supondo outra vez que há 6 valores para interpolar, $(x_1, y_1), \dots, (x_6, y_6)$, o procedimento é o seguinte:

1. Definir os polinômios de 4º grau $f_{esq.}$ e $f_{dir.}$, da mesma forma como no caso do método de Sprague;
2. Nos dois primeiros intervalos da interpolação, usar $f_{esq.}$; nos dois últimos usar $f_{dir.}$; no intervalo central usar $(f_{esq.} + f_{dir.})/2$;
3. Definir todos os pontos ξ_1, ξ_2, ξ_3 etc. onde se busca obter uma interpolação, incluindo os nodos x_1, \dots, x_6 ;
4. Para calcular uma função f preliminar, acrescentar pequenos valores d_i arbitrários aos valores obtidos em 2) em todos os pontos que não sejam nodais;
5. Calcular as diferenças finitas da 5ª ordem $\Delta^5(f, \xi_1, \xi_{i+1}, \xi_{i+2}, \xi_{i+3}, \xi_{i+4}, \xi_{i+5})$, seguindo a lógica de (18.12.a-c) e somar os seus quadrados para $i=1$ até $i=n-5$.
6. Usando o Solver de EXCEL ou algum outro recurso, minimizar a expressão calculada em 5) pela escolha otimizada dos valores de d_i .

Na sua formulação geral, como descrita acima, este método pode ser difícil de aplicar porque exige a otimização de parâmetros d_i para todos os pontos interpolados. Hoje em dia, com a ajuda de EXCEL, essa otimização não está além das possibilidades do usuário médio, mas na época em que Beers formulou o método, isso era um obstáculo formidável. Entretanto, Beers desenvolveu uma fórmula explícita para o caso mais comum onde:

1. Todos os intervalos de interpolação têm a mesma amplitude $5a$; e
2. Os pontos onde se busca uma interpolação, além dos nodos, são $x_i + a, x_i + 2a, x_i + 3a$ e $x_i + 4a$, onde x_i é um ponto nodal e $x_i + 5a = x_{i+1}$ é outro, e de forma análoga nos outros intervalos.

Neste caso, uma solução aproximada que funciona bastante bem fora dos primeiros dois e dos últimos dois intervalos é a seguinte:

$$f(x_3 + a) = (f_{esq.}(x_3+a) + f_{dir.}(x_3+a))/2 - 0,0045 \Delta(5) a^5 \quad (18.16.a)$$

$$f(x_3 + 2a) = (f_{esq.}(x_3+2a) + f_{dir.}(x_3+2a))/2 - 0,0025 \Delta(5) a^5 \quad (18.16.b)$$

$$f(x_3 + 3a) = (f_{esq.}(x_3+3a) + f_{dir.}(x_3+3a))/2 + 0,0025 \Delta(5) a^5 \quad (18.16.c)$$

$$f(x_3 + 4a) = (f_{esq.}(x_3+4a) + f_{dir.}(x_3+4a))/2 + 0,0045 \Delta(5) a^5 \quad (18.16.d)$$

onde $\Delta(5)$ pode ser calculado de diferentes formas, mas a mais simples provavelmente é

$$\Delta(5) a^5 = f_{dir.}(x_1) - f_{esq.}(x_1) \quad (18.17)$$

Nos primeiros dois intervalos, a fórmula aproximada recomendada por Beers é:

$$f(x_1 + a) = f_{esq.}(x_1+a) + 0,0183 \Delta(5) a^5 \quad (18.18.a)$$

$$f(x_1 + 2a) = f_{esq.}(x_1+2a) + 0,0177 \Delta(5) a^5 \quad (18.18.b)$$

$$f(x_1 + 3a) = f_{esq.}(x_1+3a) + 0,0102 \Delta(5) a^5 \quad (18.18.c)$$

$$f(x_1 + 4a) = f_{esq.}(x_1+4a) + 0,0033 \Delta(5) a^5 \quad (18.18.d)$$

$$f(x_2 + a) = f_{esq.}(x_2+a) + 0,0001 \Delta(5) a^5 \quad (18.18.e)$$

$$f(x_2 + 2a) = f_{esq.}(x_2+2a) + 0,0013 \Delta(5) a^5 \quad (18.18.f)$$

$$f(x_2 + 3a) = f_{esq.}(x_2+3a) + 0,0015 \Delta(5) a^5 \quad (18.18.g)$$

$$f(x_2 + 4a) = f_{esq.}(x_2+4a) + 0,0004 \Delta(5) a^5 \quad (18.18.h)$$

Nos dois últimos intervalos, as fórmulas são a imagem de espelho de (18.18.a-h), com $f_{dir.}$ em vez de $f_{esq.}$ e sinais negativos em vez de positivos.

Liu et al. (2011) experimentaram dez métodos distintos para a interpolação de curvas de fecundidade e concluíram que o método de Beers foi o que melhor funcionou para desagregar taxas quinquenais em taxas por idades simples. Isso não é de estranhar, já que o método de Beers foi desenvolvido especificamente para este propósito. O método de Sprague teve resultados quase tão bons quanto os de Beers, mas é mais versátil e tem algumas vantagens para a interpolação contínua. O método de Beers produz valores interpolados num número limitado de pontos, mas

não produz uma função contínua de interpolação que possa ser diferenciada ou integrada. Mas há situações onde é preciso fazer isso, por exemplo para derivar uma densidade a partir de uma função acumulada. Estritamente falando, a interpolação de Beers não permite isso. Entretanto, as seguintes fórmulas fornecem uma interpolação contínua para o intervalo central, da mesma forma que (8.13.a-f) no caso do método de Sprague, para permitir que pelo menos o intervalo central possa ser diferenciado, quando há necessidade.

$$f_1(x) = (- 50 (x/a)^5 + 133 (x/a)^4 - 88 (x/a)^3 - 25 (x/a)^2 + 30 (x/a)) / 384 \quad (18.19.a)$$

$$f_2(x) = (250 (x/a)^5 - 649 (x/a)^4 + 344 (x/a)^3 + 301 (x/a)^2 - 246 (x/a)) / 384 \quad (18.19.b)$$

$$f_3(x) = (- 250 (x/a)^5 + 633 (x/a)^4 - 280 (x/a)^3 - 285 (x/a)^2 - 10 (x/a) + 192) / 192 \quad (18.19.c)$$

$$f_4(x) = (250 (x/a)^5 - 617 (x/a)^4 + 248 (x/a)^3 + 173 (x/a)^2 + 138 (x/a)) / 192 \quad (18.19.d)$$

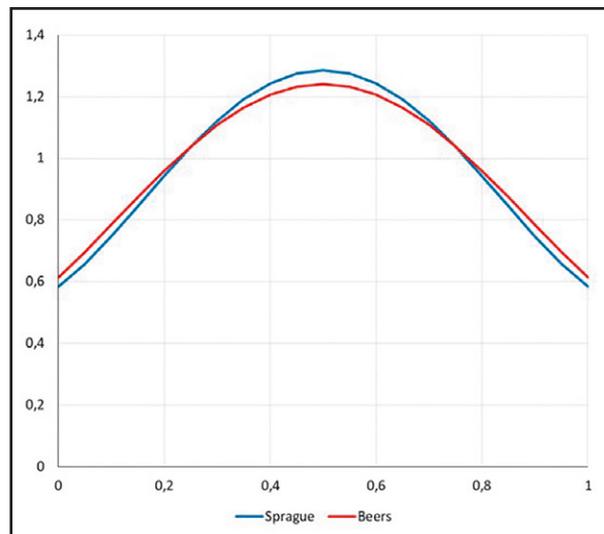
$$f_5(x) = (- 250 (x/a)^5 + 601 (x/a)^4 - 248 (x/a)^3 - 61 (x/a)^2 - 42 (x/a)) / 384 \quad (18.19.e)$$

$$f_6(x) = (50 (x/a)^5 - 117 (x/a)^4 + 56 (x/a)^3 + 9 (x/a)^2 + 2 (x/a)) / 384 \quad (18.19.f)$$

Especificamente, o equivalente da fórmula (18.14) para o caso do método de Beers é

$$\frac{d}{dx} f_4(x) + f_5(x) + f_6(x) = (1500 x^4 - 3000 x^3 + 912 x^2 + 588 x + 236) / 384 \quad (18.20)$$

Gráfico 18.4: As curvas de Sprague e Beers que descrevem um volume de 1 no intervalo central comparadas



Fonte: Elaboração própria.

As explicações sobre os métodos de Sprague e Beers providenciadas nesta seção e na anterior vão além daquelas que costumam aparecer nos textos de demografia mais antigos. Tanto o método de Sprague como o de Beers, no seu padrão (intervalos quinquenais com valores interpolados nas idades simples), podem ser tabulados com relativa facilidade. As tabelas podem ser encontradas em Siegel e Swanson (2004: Table C.14 e Table C.15). Entretanto, estas tabelas têm três limitações:

1. Elas funcionam apenas no caso padrão e não indicam como fazer a interpolação no caso de intervalos desiguais ou pontos de interpolação que não sejam os usuais.
2. Elas não permitem acomodar condições especiais, por exemplo, para evitar que a curva interpolada assumira valores negativos.
3. Como elas não mostram a forma funcional completa, elas não podem ser diferenciadas ou integradas com facilidade.

Antigamente, os cálculos envolvidos na construção das tabelas eram tão difíceis que valia a pena apresentá-las de forma padronizada, mas hoje em dia isso não é mais o caso. Espera-se que, ao mostrar a lógica subjacente, estas seções tenham estimulado os leitores a fazer seus próprios programas em EXCEL ou “R”, superando assim as limitações assinaladas acima.

18.2.3 Splines

Não existe uma diferença fundamental entre a interpolação osculatória e a interpolação baseada em splines. A diferença está mais nas origens e no desenvolvimento matemático dos métodos. A interpolação osculatória foi desenvolvida em grande medida por atuários, como um recurso pragmático para interpolar e graduar dados de mortalidade. Os splines provêm da área de engenharia³ e tiveram um desenvolvimento matemático muito mais amplo e sistemático. Dependendo das aplicações, existe uma grande variedade de splines que se diferenciam pela suas formas funcionais (polinomiais de diferentes graus, racionais, exponenciais), a sua forma de aplicação (para fins de interpolação ou aproximação) e as restrições impostas para garantir que satisfaçam certas condições na margem ou certos comportamentos globais, como a monotonicidade. Aqui a discussão será restrita a splines polinomiais do primeiro, segundo ou terceiro grau que são os mais usados. Os splines polinomiais do terceiro grau, em particular, são considerados muito úteis por terem um grau suficientemente alto para auferir certa flexibilidade, mas ao mesmo tempo suficientemente baixo para não oscilar excessivamente. O uso de splines polinomiais do quarto ou quinto grau é mais raro, embora às vezes se torne necessário devido à quantidade de restrições nos pontos extremos (por exemplo, McNeil, Trussell e Turner, 1977). Aqui a discussão se limitará ao spline natural, os splines de Hermite, os B-splines, os splines de Bézier – todos cúbicos – e os splines quadráticos de Schmertmann.

³ A palavra “spline” se refere a uma barra de metal flexível que antigamente se usava na engenharia naval para ajudar a desenhar linhas curvas que tinham que passar por um determinado ponto sob um determinado ângulo.

Embora a ideia que rege a lógica da interpolação com splines seja relativamente fácil de entender, a sua aplicação tem uma série de complicações práticas, algumas das quais serão discutidas aqui, mas não de forma exaustiva. Trata-se dos seguintes problemas:

- Quase toda a teoria para a interpolação com splines foi desenvolvida para a interpolação (ou aproximação) de valores pontuais; a interpolação de volumes ou taxas não é bem desenvolvida, mas em aplicações demográficas frequentemente precisa-se deste tipo de técnica.
- Embora os splines tendam a oscilar muito menos do que os polinômios de alto grau, se não forem bem calibrados ainda podem manter a tendência a oscilar excessivamente, principalmente se a função subjacente muda de tendência bruscamente, e boa parte da escolha de um bom método de interpolação visa controlar essa tendência.
- É preciso garantir que os valores interpolados sejam plausíveis nos extremos.
- Em aplicações demográficas quase sempre é preciso garantir que todos os valores interpolados sejam positivos e em alguns casos também é preciso garantir a monotonicidade.

A estabilidade numérica dos métodos de cálculo dos splines pode ser problemática em certos casos, de modo que torna-se necessário adotar procedimentos que evitam a extrema volatilidade dos cálculos.

Em alguns casos, em que todas estas complicações se apresentam ao mesmo tempo, a interpolação com splines pode ser tecnicamente complexa.

18.2.3.1 Spline cúbico natural

Este é um dos splines mais usados, por ser relativamente fácil de aplicar e por possuir propriedades matemáticas desejáveis. Supondo uma sequência de n pontos $(x_1, y_1), \dots, (x_n, y_n)$, o spline cúbico natural consiste de $n-1$ segmentos de polinômios cúbicos f_1, \dots, f_{n-1} definidos de tal forma que

- O polinômio cúbico f_i interpola entre o ponto (x_i, y_i) e o ponto (x_{i+1}, y_{i+1}) ;
- $f_i(x_i) = y_i$ e $f_i(x_{i+1}) = y_{i+1}$, ou seja, cada polinômio f_i reproduz os valores correspondendo aos extremos do seu intervalo;
- $f_i'(x_{i+1}) = f_{i+1}'(x_{i+1})$ e $f_i''(x_{i+1}) = f_{i+1}''(x_{i+1})$ para $i=1, \dots, n-2$, ou seja, polinômios de intervalos vizinhos têm a mesma derivada e a mesma segunda derivada no ponto de junção; e
- $f_1''(x_1) = f_{n-1}''(x_n) = 0$.

Essas condições determinam completamente quais são os coeficientes de cada polinômio f_1, \dots, f_{n-1} . O spline cúbico natural tem a propriedade desejável de ser *a mais suave* entre todas as funções f que são duas vezes diferenciáveis entre x_1 e x_n , no sentido de que minimiza a expressão I_f em (18.8).

Existem diferentes maneiras para encontrar os coeficientes de todos os $n-1$ polinômios. Uma maneira intuitiva é fixar o valor de $f_1'(x_1)$ numa estimativa inicial e – com as condições de que $f_1(x_1) = y_1$, $f_1(x_2) = y_2$ e $f_1''(x_1) = 0$, determinar os quatro coeficientes de f_1 . Esses valores determinam $f_2'(x_2)$ e $f_2''(x_2)$ e com a condição adicional de que $f_2(x_2) = y_2$ e $f_2(x_3) = y_3$ é possível determinar todos os coeficientes de f_2 . Continuando desta forma podem ser determinados todos os coeficientes de todos os polinômios até f_n e o valor de $f_n''(x_n)$ pode ser calculado. Finalmente se escolhe o parâmetro inicial $f_1'(x_1)$ de tal forma que $f_n''(x_n) = 0$. É assim que o Gráfico 18.5 foi gerado, por exemplo. Além de ser mais intuitivo, esse procedimento tem a vantagem de ser mais fácil de adaptar à interpolação de taxas (ver seção 18.4). Entretanto, o método nem sempre é numericamente estável. Um método mais seguro procura estimar os valores de $f_1''(x_1)$, ..., $f_n''(x_n)$ e a partir disso obter a função inteira de interpolação pela integração. Isso implica a inversão da seguinte matriz:

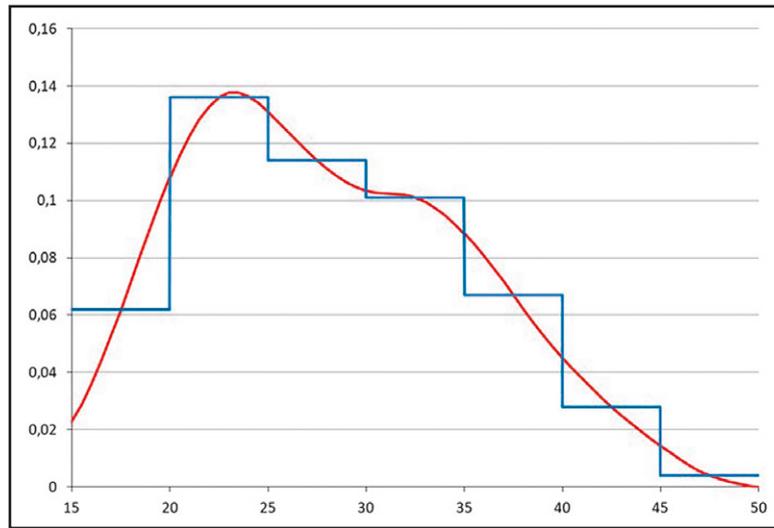
$$\begin{bmatrix} 2(x_3 - x_1) & x_3 - x_2 & 0 & 0 & 0 & 0 & 0 \\ x_3 - x_2 & 2(x_4 - x_2) & x_4 - x_3 & 0 & 0 & 0 & 0 \\ 0 & x_4 - x_3 & 2(x_5 - x_3) & x_5 - x_4 & 0 & 0 & 0 \\ 0 & 0 & \dots\dots & \dots\dots & \dots\dots & 0 & 0 \\ 0 & 0 & 0 & \dots\dots & \dots\dots & \dots\dots & 0 \\ 0 & 0 & 0 & 0 & x_{n-2} - x_{n-3} & 2(x_{n-1} - x_{n-3}) & x_{n-1} - x_{n-2} \\ 0 & 0 & 0 & 0 & 0 & x_{n-1} - x_{n-2} & 2(x_n - x_{n-2}) \end{bmatrix} \quad (18.21)$$

No caso mais comum de pontos equidistantes com separações de 5 anos, a diagonal é igual a 10 e as subdiagonais a 5. A inversa de (18.21) precisa ser multiplicada pelo seguinte vetor de n por 1:

$$\begin{bmatrix} 6((y_3 - y_2)/(x_3 - x_2) - (y_2 - y_1)/(x_2 - x_1)) \\ 6((y_4 - y_3)/(x_4 - x_3) - (y_3 - y_2)/(x_3 - x_2)) \\ \dots\dots \\ \dots\dots \\ \dots\dots \\ \dots\dots \\ 6((y_n - y_{n-1})/(x_n - x_{n-1}) - (y_{n-1} - y_{n-2})/(x_{n-1} - x_{n-2})) \end{bmatrix} \quad (18.22)$$

O resultado fornece os valores de $f_2''(x_2)$, ..., $f_{n-1}''(x_{n-1})$. As outras duas segundas derivadas são $f_1''(x_1) = 0$ e $f_n''(x_n) = 0$. Os detalhes da derivação de f_1' , ..., f_{n-1} a partir de f_1'' , ..., f_n'' não serão mostrados aqui.

Gráfico 18.5: Interpolação da curva de fecundidade de Cabo Verde (2010) com um spline cúbico natural



Fonte: Censo Demográfico de Cabo Verde (2010).

O Gráfico 18.5 mostra uma interpolação da curva de fecundidade de Cabo Verde calculada a partir do Censo de 2010, baseada no spline cúbico natural. O gráfico em forma de escada mostra as TEFs para os sete intervalos quinquenais que são interpoladas pela linha suave, usando o critério de que o valor da curva no ponto médio dos sete intervalos (17,5, ..., 47,5) deve ser igual à TEF daquele intervalo. Em realidade, este não é o melhor critério possível; na seção 18.4 será introduzida uma alternativa melhor. Além dos sete pontos que definem as TEFs nos intervalos, incluiu-se um oitavo ponto (50,0), para finalizar a curva de forma plausível. Portanto, as condições que definem a curva do Gráfico 18.5 são:

- $f_1(17,5) = TEF(15-19) = 0,062$
- $f_1(22,5) = f_2(22,5) = TEF(20-24) = 0,136$
-
- $f_6(47,5) = f_7(47,5) = TEF(45-49) = 0,004$
- $f_1'(22,5) = f_2'(22,5)$
-
- $f_6'(47,5) = f_7'(47,5)$
- $f_1''(17,5) = 0$
- $f_1''(22,5) = f_2''(22,5)$
-
- $f_6''(47,5) = f_7''(47,5)$
- $f_7''(50) = f_7(50) = 0$

18.2.3.2 Spline cúbico de Hermite

O spline de Hermite, também chamado C-spline, tem uma forma ainda mais simples do que o spline natural, mas em vez de ser duas vezes diferenciável, geralmente é diferenciável apenas uma vez⁴. A ideia básica é usar a seguinte expressão para f_i :

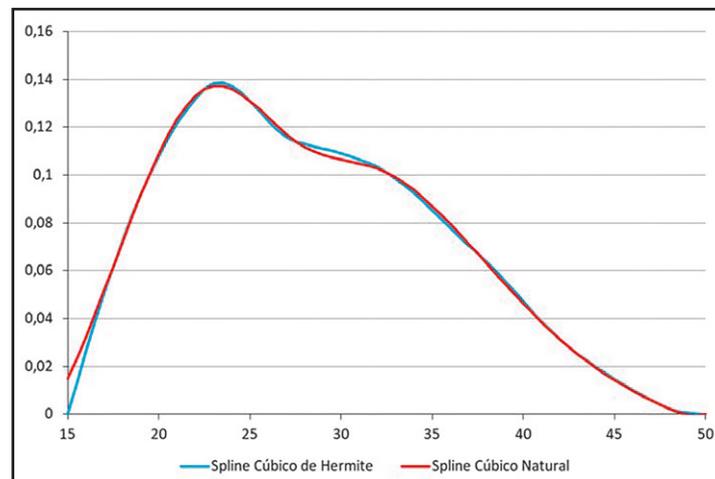
$$f_i(x) = (2t^3 - 3t^2 + 1)y_i + (t^3 - 2t^2 + t)(x_{i+1} - x_i)f'_i(x_i) + (-2t^3 + 3t^2)y_{i+1} + (t^3 - t^2)(x_{i+1} - x_i)f'_i(x_{i+1}) \quad (18.23)$$

onde $t = (x - x_i)/(x_{i+1} - x_i)$. Alternativamente, se pode especificar $f_i(x)$ da seguinte forma:

$$f_i(x) = (2y_i - 2y_{i+1} + f'_i(x_i) + f'_i(x_{i+1}))x^3 + (-3y_i + 3y_{i+1} - 2f'_i(x_i) - f'_i(x_{i+1}))x^2 + f'_i(x_i)x + y_i \quad (18.24)$$

Para que a expressão seja diferenciável em x_{i+1} , é preciso que $f'_i(x_{i+1}) = f'_{i+1}(x_{i+1})$ para $i=1, \dots, n-2$. A não ser que haja alguma condição adicional para ser satisfeita, as derivadas em todos os n pontos observados são parâmetros livres que devem ser escolhidos para minimizar a irregularidade da função resultante. Em vez de usar o critério (18.8), que exige que a função seja duas vezes diferenciável, é melhor usar o comprimento total do arco da linha interpolada, como em (18.10) ou (18.11). O Gráfico 18.6 compara os resultados deste procedimento para interpolar a fecundidade de Cabo Verde em 2010 com os resultados obtidos aplicando o spline cúbico natural, usando os mesmos oito pontos. A forma das duas curvas é muito semelhante. A única diferença mais aparente é que o spline de Hermite é zero aos 15 anos, mas esta foi uma condição adicional imposta ao processo de ajuste; não é uma característica intrínseca da curva de Hermite como tal.

Gráfico 18.6: Interpolação da curva de fecundidade de Cabo Verde (2010) com um spline cúbico natural e com o spline cúbico de Hermite



Fonte: Censo Demográfico de Cabo Verde (2010).

⁴ É possível fazer a expressão (18.23) ou (18.24) duas vezes continuamente diferenciável, mas isso requer condições adicionais em cima dos valores das derivadas $f'_i(x_i)$ nos nodos. Se a curva for duas vezes continuamente diferenciável em todos os nodos, com segunda derivada igual a zero nos extremos, ela se torna um spline natural.

18.2.3.3 B-splines

B-splines são funções simétricas que são iguais a zero, exceto na sua região de suporte que varia com o seu grau. O B-spline quadrático é uma vez diferenciável e tem um suporte de 3, o B-spline cúbico é duas vezes diferenciável e tem um suporte de 4 etc. As fórmulas dos B-splines mais comuns são as seguintes:

B-spline Linear

$$B_{1,1} = x \quad (0 \leq x \leq 1) \quad (18.25.a)$$

$$B_{2,1} = 1 - x \quad (1 \leq x \leq 2) \quad (18.25.b)$$

B-spline Quadrático

$$B_{1,2} = x^2 / 2 \quad (0 \leq x \leq 1) \quad (18.26.a)$$

$$B_{2,2} = (-2x^2 + 6x - 3) / 2 \quad (1 \leq x \leq 2) \quad (18.26.b)$$

$$B_{3,2} = (3 - x)^2 / 2 \quad (2 \leq x \leq 3) \quad (18.26.c)$$

B-spline Cúbico

$$B_{1,3} = x^3 / 6 \quad (0 \leq x \leq 1) \quad (18.27.a)$$

$$B_{2,3} = (-3x^3 + 12x^2 - 12x + 4) / 6 \quad (1 \leq x \leq 2) \quad (18.27.b)$$

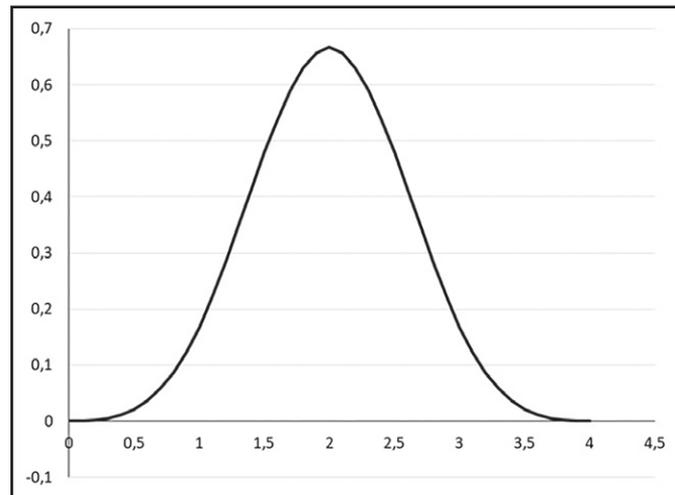
$$B_{3,3} = (-3(4-x)^3 + 12(4-x)^2 - 12(4-x) + 4) / 6 \quad (2 \leq x \leq 3) \quad (18.27.c)$$

$$B_{4,3} = (4 - x)^3 / 6 \quad (3 \leq x \leq 4) \quad (18.27.d)$$

É de notar que (18.27.a-d) define uma densidade de probabilidade, com um volume total (a integral de 0 a 4) igual a 1. A curva pode ser deslocada horizontalmente ou expandida ou contraída, aumentando ou reduzindo a sua amplitude ou variância:

$$B_{i,3,m,k} = B_{i,3}((x - m) / k) / k \quad (18.28)$$

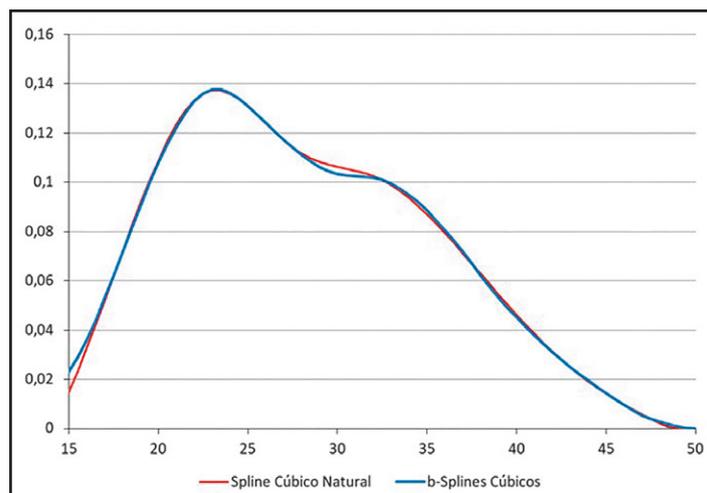
Inclusive é possível generalizar as fórmulas acima para o caso de intervalos desiguais, o que resulta em funções assimétricas, mas aqui serão tratados apenas os B-splines simétricos.

Gráfico 18.7: Representação gráfica do B-spline cúbico na sua forma padrão ($m = 0$; $k = 1$)

Fonte: Elaboração própria

A ideia básica do uso de B-splines é que funções que possuem certa regularidade podem ser aproximadas por um conjunto de B-splines. No caso em que não se conhece a função em todo o seu domínio, mas apenas em alguns pontos, os B-splines podem ser usados para graduar estes pontos. Embora os B-splines não sejam propriamente instrumentos para a interpolação, muitas vezes é possível conseguir uma aproximação tão boa que o resultado acaba sendo esse, até mesmo com um número relativamente pequeno de B-splines. O Gráfico 18.8 mostra uma interpolação da curva de fecundidade de Cabo Verde calculada a partir do Censo de 2010. A curva de interpolação é a soma ponderada de quatro B-splines, todos com uma amplitude de 28 anos (ou seja, $k = 7$) e localizados nos pontos $m_1 = 8,85$; $m_2 = 17,91$; $m_3 = 25,11$ e $m_4 = 33,84$, com pesos de 1,419, 0,991, 0,213 e -0,046, respectivamente. É de notar que a soma desses quatro pesos é igual à TFT: $1,419 + 0,991 + 0,213 - 0,046 = 2,577$.

Gráfico 18.8: Interpolação da curva de fecundidade de Cabo Verde (2010) com o spline cúbico natural e com uma combinação de quatro B-splines cúbicos



Fonte: Censo Demográfico de Cabo Verde (2010).

A dificuldade da aplicação deste método de interpolação reside em encontrar os quatro nodos m_1, m_2, m_3 e m_4 . Esse posicionamento é crítico e infelizmente não existem regras claras para orientar como deve ser feito. Se os pontos não forem escolhidos corretamente, o resultado pode ser uma curva bastante irregular que reproduz as sete TEFs, mas que oscila violentamente. Para produzir o Gráfico 18.9, se minimizou a irregularidade da função interpolada por meio do recurso Solver de EXCEL, variando os valores de m_1, m_2, m_3 e m_4 . O critério de regularidade usado foi o comprimento total do arco da linha interpolada entre 15 e 50 anos. Como se vê, os resultados são muito próximos dos resultados da interpolação com um spline cúbico natural com o critério de regularidade (18.8).

18.2.3.4 Curvas de Bézier

As curvas de Bézier são relativamente pouco usadas na demografia, mas elas são o modo de interpolação preferido no mundo dos gráficos de computador. A curva de interpolação que o EXCEL usa para implementar os gráficos de dispersão com linhas suaves é quase idêntica a uma curva de Bézier de terceiro grau⁵. Entretanto, o EXCEL apenas mostra as curvas graficamente e não fornece informação sobre a sua forma matemática. Vale a pena conhecer o método por detrás desta interpolação porque ela é bastante flexível e permite interpolações que tendem a ser mais suaves do que as mostradas até agora.

A curva de Bézier do terceiro grau tem uma certa semelhança com o spline natural e com o spline de Hermite. A diferença principal reside no fato de que a curva de Bézier não se limita a funções convencionais onde cada x está associado a apenas um valor de y , mas interpola no plano (x,y) inteiro, podendo produzir resultados que não são funções, no sentido estrito da palavra. Por exemplo, uma curva de Bézier pode ser parecida com um círculo, onde (quase) todos os valores de x estão associados a *dois* valores de y . Para este fim, o algoritmo, além de usar os pontos de interpolação $(x_1, y_1), \dots, (x_n, y_n)$ propriamente ditos, usa *pontos de controle*, que não fazem parte da curva mas que atraem a curva em determinadas direções. No caso da curva de Bézier de terceiro grau, cada segmento de interpolação tem dois pontos de controle, além dos dois pontos por interpolar. Sendo assim, a fórmula da curva é a seguinte:

$$B_i(t) = (1-t)^3 P_{i0} + 3(1-t)^2 t P_{i1} + 3(1-t) t^2 P_{i2} + t^3 P_{i3} \quad (18.29)$$

onde $P_{i0} = (x_i, y_i)$ e $P_{i3} = (x_{i+1}, y_{i+1})$ são os pontos de interpolação nos extremos do intervalo e P_{i1} e P_{i2} são pontos de controle no plano (x,y) . Como na interpolação de Hermite em (18.23), o parâmetro t varia entre 0 e 1, dependendo da posição do ponto interpolado, mais próximo de P_{i0} ou de P_{i3} . O desafio agora consiste em encontrar as posições ótimas para os dois pontos de controle.

O método para este fim consiste nos seguintes passos. Primeiro se define uma matriz parecida com (18.21), mas com uma estrutura ainda mais simples e um número de filas e colunas que corresponde ao número de intervalos a interpolar (n):

⁵ Em realidade, o EXCEL usa uma adaptação do spline de Catmull-Rom que, em casos onde o espaçamento dos pontos de interpolação é razoavelmente regular, é praticamente igual a uma curva de Bézier.

$$\begin{bmatrix}
 2 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 4 & 1 & 0 & 0 & 0 & 0 \\
 0 & 1 & 4 & 1 & 0 & 0 & 0 \\
 0 & 0 & \dots & \dots & \dots & 0 & 0 \\
 0 & 0 & 0 & \dots & \dots & \dots & 0 \\
 0 & 0 & 0 & 0 & 1 & 4 & 1 \\
 0 & 0 & 0 & 0 & 0 & 2 & 7
 \end{bmatrix} \quad (18.30)$$

Esta matriz é invertida e o resultado multiplicado pela seguinte matriz de n por 2:

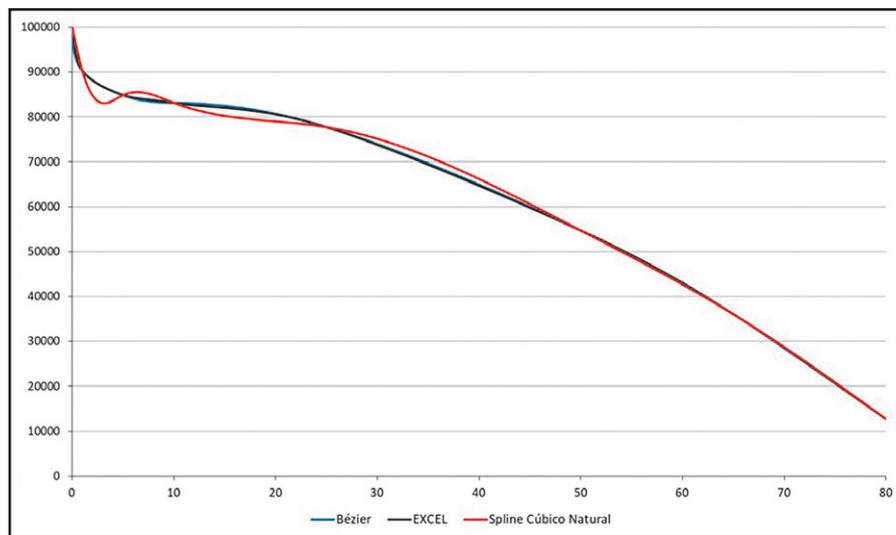
$$\begin{bmatrix}
 x_1 + 2x_2 & y_1 + 2y_2 \\
 4x_2 + 2x_3 & 4y_2 + 2y_3 \\
 4x_3 + 2x_4 & 4y_3 + 2y_4 \\
 \dots & \dots \\
 \dots & \dots \\
 4x_{n-1} + 2x_n & 4y_{n-1} + 2y_n \\
 8x_n + x_{n+1} & 8y_n + y_{n+1}
 \end{bmatrix} \quad (18.31)$$

O resultado desta multiplicação matricial são as coordenadas dos n pontos de controle P_{11}, \dots, P_{n1} , ou seja, $(x_{11}, y_{11}), \dots, (x_{n1}, y_{n1})$. Os segundos pontos de controle P_{12}, \dots, P_{n2} são calculados como

$$\begin{bmatrix}
 x_1 + 2x_2 & y_1 + 2y_2 \\
 4x_2 + 2x_3 & 4y_2 + 2y_3 \\
 4x_3 + 2x_4 & 4y_3 + 2y_4 \\
 \dots & \dots \\
 \dots & \dots \\
 4x_{n-1} + 2x_n & 4y_{n-1} + 2y_n \\
 8x_n + x_{n+1} & 8y_n + y_{n+1}
 \end{bmatrix} \quad (18.32)$$

Agora existem todos os elementos para poder calcular (18.29). É de notar que a variável que parametriza (18.29) é $t = (x - x_i)/(x_{i+1} - x_i)$, o que significa que não se pode escolher a coordenada x do ponto interpolado diretamente. Da mesma forma que y , o ponto x é o resultado de (18.29) para um determinado valor de t . Se for preciso avaliar (18.29) para um determinado valor de x , primeiro é preciso estabelecer o valor de t que produz esse valor de x , para depois calcular o valor correspondente de y .

Gráfico 18.9: Interpolação da função ℓ_x da tábua de vida masculina de Moçambique (2007) baseada nas idades de 0, 1, 5, 10, 25, 50, 65 e 80 anos usando o spline de Bézier, o procedimento de EXCEL e o spline cúbico natural



Fonte: Pontos da curva ℓ_x conforme a tábua de vida produzida pelo INE, Moçambique.

O Gráfico 18.9 mostra a interpolação da curva ℓ_x da tábua de vida masculina de Moçambique elaborada pelo INE a partir dos dados do Censo de 2007. Os pontos interpolados são as idades de 0, 1, 5, 10, 25, 50, 65 e 80 anos. Como se percebe, a curva obtida pelo método de Bézier é muito suave e se comporta de forma plausível diante da mudança de tendência que ocorre entre as idades de 1 e 5 anos. A curva também reproduz de forma muito próxima à interpolação feita pelo EXCEL. O spline cúbico natural, por outro lado, oscila nestas idades, o que reduz significativamente a sua utilidade como função de interpolação⁶. Apesar desta desvantagem, Mina-Valdés (2011) conseguiu construir uma função ℓ_x satisfatória para o caso do México usando splines naturais para quatro intervalos (0-5, 6-24, 24-69 e 70-100). A desvantagem da curva de Bézier é que ela funciona intrinsecamente com valores pontuais e não se adapta facilmente à interpolação de volumes ou taxas. Para interpolar taxas é preciso integrar a função de interpolação, o que no caso da função de Bézier é mais difícil do que nas funções apresentadas anteriormente.

⁶ Uma alternativa para a função de Bézier que tem propriedades de estabilidade semelhantes é um spline natural ponderado por $1/(1+(f'_i(x))^2)^{2.5}$, que corrige a tendência de (18.8) a exagerar a curvatura em situações onde esta aumenta ou diminui muito rapidamente. Essa ponderação reduz a tendência à oscilação (Šalkauskas, 1984). Os detalhes do método não serão mostrados aqui.

18.2.3.5 Os splines calibrados de Schmertmann

Os splines calibrados (CS, de “calibrated splines”, em inglês) de Schmertmann (2014), para a análise da fecundidade por idade, não são propriamente um método de interpolação, pois não reproduzem os dados originais exatamente. Em vez disso aplicam alguma suavização. Mas como o método é parecido com o de Beers e envolve o uso de B-splines, é tratado aqui. O método dá bons resultados no ajuste de curvas de fecundidade a partir de dados anuais ou quinquenais (Grigoriev e Jdanov, 2015). As diferenças principais com o método de Beers são as seguintes:

1. Enquanto a curva de Beers é um polinômio de quinto grau, os CS são combinações de B-splines de segundo grau, do tipo descrito em (18.26.a-c), embora os seus nodos tenham uma distância de 2 anos de idade, em vez de 1.
2. Em vez de usar um critério matemático de suavidade, como a minimização da soma de quadrados das diferenças de quinto grau usada por Beers, o método CS é calibrado com dados observados de fecundidade por idade simples da Human Fertility Database (HFD) e da International Data Base (IDB), do US Bureau of the Census.
3. O método CS toma em conta o tamanho da amostra da qual os dados observados provêm. Em amostras grandes, dá mais peso às observações; em amostras pequenas dá mais peso aos padrões teoricamente esperados.

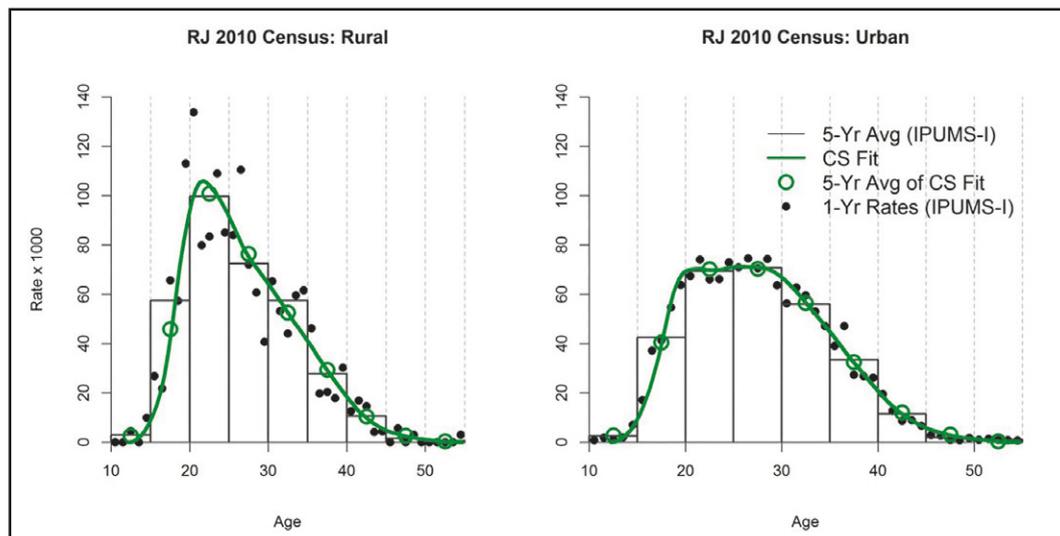
O método é uma variante dos chamados P-splines (Eilers e Marx, 1996), onde P indica “penalidade”. Trata-se de combinações de B-splines quadráticos cujos coeficientes são determinados para minimizar as penalidades decorrentes de desvios em relação aos dados observados e aos perfis etários identificados na HFD e IDB. Em princípio o método pode ser aplicado a qualquer tipo de intervalo: intervalos iguais, desiguais, sobrepostos (por exemplo, 15-24 e 20-29) e incompletos (por exemplo, apenas definidos até os 35 anos). No caso geral, os cálculos podem ser complexos, mas nos casos típicos de 7 intervalos (15-19, , 45-49) ou 9 intervalos iguais (10-14, , 50-54) o procedimento resultante é surpreendentemente simples e consiste simplesmente em somar as taxas de fecundidade observadas em cada intervalo multiplicadas por uma série de coeficientes cujos valores dependem da idade à qual a estimativa se refere e do tamanho da amostra (100, 1.000, 10.000 ou 100.000).

Por exemplo, para Cabo Verde em 2010 a taxa suavizada para idade exata 26,25 no modelo CS, assumindo um tamanho amostral relativamente pequeno de 1.000 por intervalo, seria

$$F(26,25) = -0,0965 \text{ TEF}(15-19) + 0,3025 \text{ TEF}(20-24) + \dots -0,0035 \text{ TEF}(45-49) \quad (18.33)$$

e há outras constantes que correspondem às idades 12,25, 12,75, , 54,75. O sítio de web <http://schmert.net/calibrated-spline/REBEP> apresenta vários documentos que ilustram o método e fornecem os valores das constantes. O Gráfico 18.10 mostra uma aplicação para a suavização das TEFs das áreas rurais e urbanas do Estado do Rio de Janeiro.

Gráfico 18.10: Perfis de splines calibrados para os residentes rurais (esquerda) e urbanos (direita) do Estado do Rio de Janeiro (2010)



Fonte: Schmertmann (2014): Graph 3.

18.3 INTERPOLAÇÃO DE VOLUMES E TAXAS

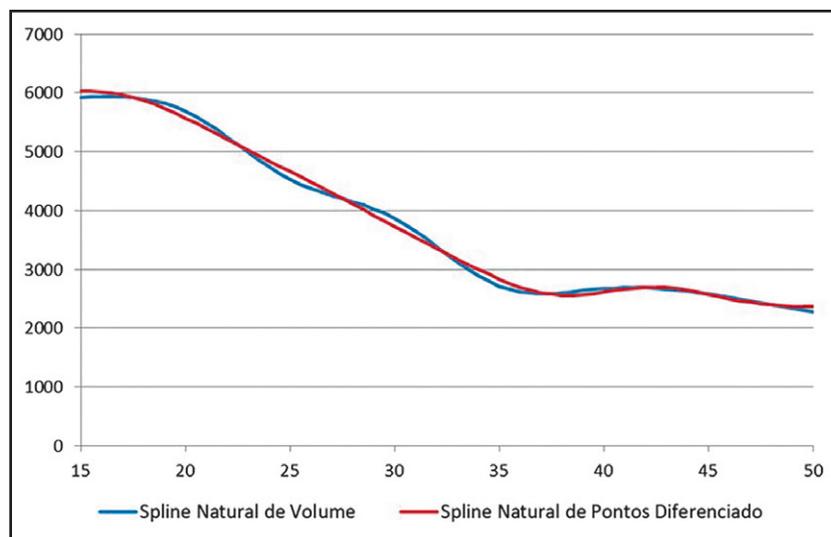
As interpolações apresentadas nos Gráficos 18.6, 18.7 e 18.8 todas foram feitas com base no critério de que no meio de cada intervalo etário a função precisava assumir o valor da TEF daquele intervalo. Este é o procedimento mais fácil e a maior parte da literatura sobre métodos de interpolação se limita a interpolações de valores pontuais. Entretanto, para a maioria das quantidades demográficas, incluindo volumes de população e taxas de diferentes tipos, não é o procedimento mais correto e não preserva a consistência dos resultados. Em muitos casos os erros introduzidos são relativamente pequenos e, como a interpolação de valores pontuais é mais fácil, pode não valer a pena aplicar um procedimento mais complexo, mas há casos em que as diferenças não podem ser ignoradas. Aqui se mostrarão alguns exemplos baseados na interpolação linear e por splines cúbicos.

No caso de volumes associados a intervalos, a solução mais comum é relativamente simples. Constrói-se primeiro uma função $f(x)$ que acumula a população (ou outro tipo de volume) ao longo de todos os intervalos relevantes. Por exemplo, se o objetivo for interpolar a população ao longo do intervalo (5,10) tomando em conta as populações em (1,5) e (10,15), a função $f(x)$ seria 0 em $x=1$, ${}_4P_1$ em $x=5$, ${}_4P_1 + {}_5P_5$ em $x=10$ e ${}_4P_1 + {}_5P_5 + {}_5P_{10}$ em $x=15$. Agora $f(x)$ pode ser interpolada da mesma forma como seria o caso com qualquer função que descreve valores pontuais em $x = 1, 5, 10$ e 15 . Uma vez tendo a forma funcional de $f(x)$, se diferencia a expressão para obter a densidade de população $f'(x)$ ao longo do intervalo. É importante dar-se conta do fato de que o grau de regularidade da função diferenciada é um a menos do que da função original. Por exemplo, se a população acumulada foi interpolada com uma função uma vez diferenciável, a densidade da população será contínua, mas não diferenciável. Se o objetivo for obter uma função diferenciável para descrever a densidade da população, a função escolhida para interpolar a população acumulada precisa ser pelo menos duas vezes diferenciável.

A linha vermelha no Gráfico 18.11 ilustra o procedimento descrito acima no caso da população feminina entre 15 e 50 anos de Cabo Verde, que é a população subjacente às TEFs interpoladas anteriormente. A população foi acumulada até 15 anos ($f(x) = 0$), 20 anos ($f(x) = {}_5P_{15}$), 25 anos ($f(x) = {}_5P_{15} + {}_5P_{20}$) etc., os pontos resultantes foram interpolados com um spline cúbico natural e a curva de interpolação foi diferenciada. A curva final é um spline quadrado uma vez continuamente diferenciável, mas como se pode ver no Gráfico 18.11, o seu comportamento é bastante suave.

A linha azul no Gráfico 18.11 foi construída diretamente como um spline cúbico natural de volumes. Isso quer dizer que as condições principais que definem a curva não são que ela precise passar por determinados pontos, mas que as suas integrais sobre os intervalos $(15,20)$, $(20,25)$, $(25,30)$ etc. tenham os valores corretos correspondendo às populações naqueles grupos etários. Além disso, a curva precisa satisfazer as condições de continuidade, a saber que seus segmentos $f_i(x)$ tenham os mesmos valores, as mesmas derivadas e as mesmas segundas derivadas nos nodos comuns (15, 20, 25 anos etc.). Finalmente, por ser um spline natural, as suas segundas derivadas em $x=15$ e $x=50$ precisam ser iguais a zero. No total, isso significa que há 20 condições por satisfazer, 1 a menos do que no caso da interpolação de valores pontuais. Esse último parâmetro pode ser escolhido de tal forma que maximize a regularidade da curva e minimize as oscilações. Mesmo assim e apesar de estar bem próxima à curva anterior, o Gráfico 18.11 sugere que a curva azul oscila um pouco mais.

Gráfico 18.11: A população feminina entre 15 e 50 anos de Cabo Verde (2010), interpolada com um spline cúbico natural de volume e com um spline cúbico natural aplicado aos valores acumulados até os pontos de 15, 20,, 50 anos e posteriormente diferenciado



Fonte: Censo Demográfico de Cabo Verde (2010).

No caso de taxas ou razões, existem dois procedimentos que dão resultados consistentes, a saber:

- Interpolando o numerador e o denominador separadamente e depois dividir os resultados. Essa foi a estratégia aplicada, por exemplo, na derivação da fórmula (17.45). O inconveniente é que o resultado acaba sendo uma razão entre as duas formas funcionais escolhidas

para interpolar o numerador e o denominador. Por um lado, esse resultado pode ser difícil de integrar analiticamente, por exemplo para obter valores para idades simples. Por outro lado, mesmo que tanto o numerador como o denominador sejam funções perfeitamente suaves e regulares, a razão entre os dois pode ter comportamentos pouco desejáveis.

- A outra possibilidade é determinar de antemão qual é a forma funcional que o resultado final deve ter (por exemplo, linear, polinômio de grau n etc.) e depois *ponderar* o procedimento pela população subjacente, para tomar em conta, por exemplo, que a TEF para o grupo etário de 15-19 anos precisa refletir que (geralmente) há mais mulheres de 15 anos do que mulheres de 19 anos. Embora este seja o procedimento preferido, também é o mais complicado do ponto de vista computacional.

Aqui será ilustrado o segundo procedimento com o exemplo de uma interpolação linear e um spline natural cúbico.

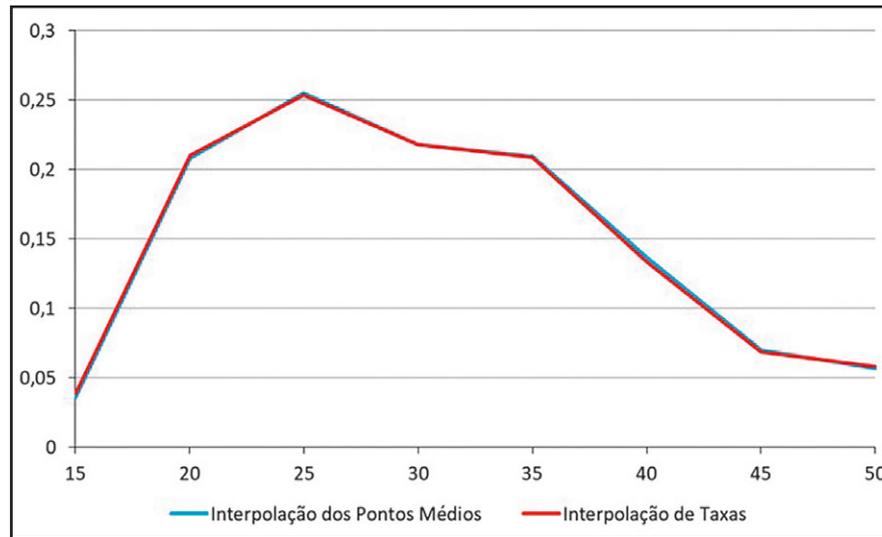
Em ambos os casos, o primeiro passo é a interpolação do padrão etário subjacente às taxas. Em princípio isso pode ser feito com qualquer método de interpolação que se julgue conveniente e não há nenhum motivo para que o mesmo deva ser coerente com o método de interpolação da taxa. Mas para fins ilustrativos se resolveu aqui interpolar a população com o mesmo tipo de curva usada para interpolar a taxa. No caso da interpolação linear, isso significa que a população subjacente foi interpolada com o procedimento descrito na Tabela 18.1, com a diferença de que os grupos etários neste caso se referem às idades reprodutivas, ou seja, 15-19, ..., 45-49. Sendo assim, a população de cada uma dessas faixas etárias agora é representada por uma função linear de ponderação $P_i(x) = A_i + B_i x$.

Para interpolar a taxa se segue um procedimento análogo ao descrito em (18.2) e (18.3). Primeiro se define uma inclinação preliminar para o segmento da primeira faixa etária. Não se pode supor mais que o valor da taxa corresponda ao valor da curva interpolada no ponto médio do intervalo etário. Em vez disso, a equação a ser satisfeita é a seguinte:

$$\int_{x_1}^{x_1+n} P_1(x) f_1(x) dx / {}_n P_{x_1} = {}_n f_{x_1} \quad (18.34)$$

Nesta expressão, $f_1(x)$ é uma função linear com uma inclinação conhecida e um nível desconhecido, enquanto todos os parâmetros de $P_1(x)$ são conhecidos. A solução do nível a partir de (18.34) é simples. Tendo a inclinação e o nível para o primeiro segmento, se conhecem os valores y_1 e y_2 do segmento interpolado em $x_1 = 15$ e $x_2 = x_1 + n = 20$. Agora o valor de y_2 pode ser usado para calcular a função $f_2(x)$ no segundo intervalo, outra vez usando (18.34), e assim adiante. Finalmente se ajusta a escolha preliminar da inclinação do primeiro segmento para minimizar as oscilações de $f(x)$ ou para evitar que assumam valores implausíveis. O Gráfico 18.12 mostra a interpolação das taxas de fecundidade do Censo de Moçambique de 2007, comparada com o resultado obtido usando o método mais simples de interpolação dos pontos médios dos intervalos, identificando cada ponto médio com a TEF daquele intervalo. Embora haja diferenças, elas são muito pequenas. A maior diferença ocorre no grupo etário de 15-19 anos, onde $f_1(15) = 0,03606$ no caso da interpolação de pontos médios e $f_1(15) = 0,03869$ no caso do procedimento baseado em (18.34).

Gráfico 18.12: Interpolação da TEF do Censo de Moçambique de 2007 usando o método de interpolação de pontos médios dos intervalos e o método de interpolação de taxas



Fonte: Censo Demográfico de Moçambique (2007).

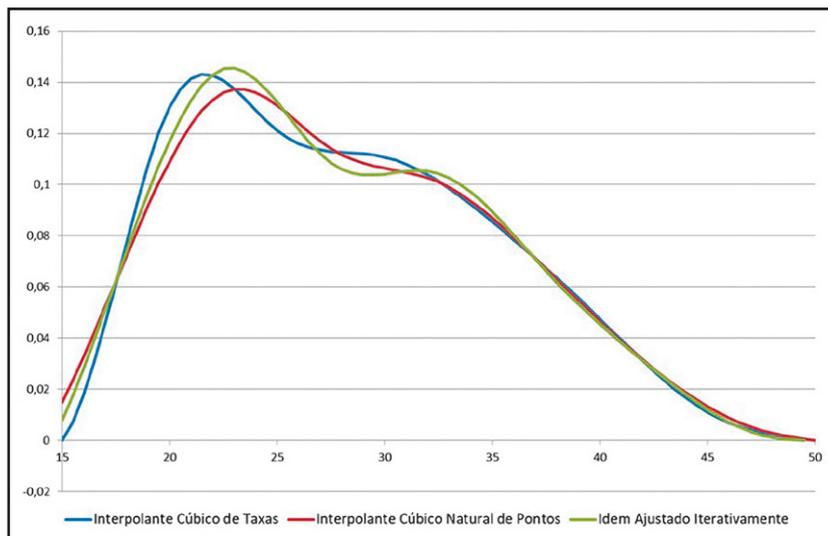
A razão principal da correspondência próxima das interpolações lineares no Gráfico 18.12 é a relativa rigidez da interpolação linear que não permite modificações fortes no perfil etário. A situação é diferente quando se usa uma função de interpolação mais flexível, como um spline cúbico. Esta situação é mostrada no Gráfico 18.12. A curva vermelha é simplesmente o spline cúbico natural do Gráfico 18.5. Embora essa curva funcione bem para interpolar os pontos médios dos intervalos, ela produz resultados inconsistentes para as taxas, quando essas são calculadas com base no perfil etário $P_i(x)$ da interpolação cúbica natural do Gráfico 18.12, em comparação com as observadas:

	Taxas Observadas	Taxas Calculadas com Base na Curva
15-19	0,062	0,0617
20-24	0,136	0,1288
25-29	0,114	0,1171
30-34	0,101	0,0995
35-39	0,067	0,0670
40-44	0,028	0,0288
45-49	0,004	0,0053

Como pode ser visto nesses dados, algumas inconsistências são significativas, particularmente na faixa de 20-24 anos. Uma primeira estratégia para resolver o problema é o ajuste iterativo dos valores observados, de tal forma que os resultados calculados sejam corretos. Por exemplo, no caso do intervalo de 20-24 anos, o valor observado de 0,136 é substituído por $0,136 \cdot 0,136 / 0,1288 = 0,1436$. E assim para os outros intervalos. Isso precisa ser feito várias vezes para obter o resultado

final correto, que é mostrado no Gráfico 18.2 como a linha verde. Embora este procedimento reproduza os valores corretos das taxas, o perfil etário, com a subida da fecundidade por volta dos 30 anos, parece implausível.

Gráfico 18.13: Interpolação da curva de fecundidade de Cabo Verde (2010) com o spline cúbico natural aplicado a valores pontuais e com um spline cúbico aplicado a taxas



Fonte: Censo Demográfico de Cabo Verde (2010).

A curva azul no Gráfico 18.13 reproduz os valores corretos das taxas e parece ter um perfil etário mais plausível do que a curva verde. Esta curva foi produzida da mesma forma que a curva do Gráfico 18.5, com a diferença de que, em vez da condição de que precisa passar pelos pontos $(17,5, TEF(15-19))$, $(22,5, TEF(20-24))$ etc., a condição agora é o equivalente de (18.34). Portanto, os parâmetros de $f_1(x)$ são determinados pela condição de que $f(15) = 0$, a equação (18.34) e dois parâmetros livres. Os parâmetros de $f_2(x)$ são determinados pela condição de que o valor e a primeira e segunda derivadas de f_2 em $x=20$ devem coincidir com os mesmos valores de f_1 , mais a equação (18.34), e assim para os intervalos seguintes. Finalmente, os dois parâmetros livres do primeiro intervalo são determinados pela condição de que $f(50) = f'(50) = 0$. O problema deste procedimento é que tende a ser numericamente instável. Para conseguir estimar a curva no Gráfico 18.13 foi necessário começar o procedimento aos 30 anos, expandindo-o para ambos os lados, em vez da alternativa mais intuitiva que seria começar aos 15 anos e avançar para as idades mais avançadas.

18.4 VALORES NEGATIVOS E TENDÊNCIAS NÃO ADMISSÍVEIS

Com poucas exceções (o saldo migratório, por exemplo), as quantidades que se manejam na demografia são inerentemente não negativas. Consequentemente é preciso garantir que as interpolações destas quantidades não assumam valores negativos, mas se não são tomadas precauções especiais, valores negativos podem surgir com relativa facilidade. Não existem métodos gerais para evitar isso: as medidas a tomar variam segundo as circunstâncias e a causa do problema. A seguinte lista apenas apresenta algumas sugestões de medidas possíveis:

- Às vezes, valores negativos surgem como consequência das oscilações excessivas na função interpolada. Quando isso ocorre, é pertinente perguntar se uma interpolação de outro tipo poderia reduzir as oscilações e evitar os valores negativos. Por exemplo, em vez de usar um spline natural, seria possível considerar uma curva de Bézier que normalmente é mais estável.
- Se o valor negativo surge num dos extremos, às vezes é possível impor condições adicionais ao cálculo da curva, forçando-a a assumir valores plausíveis (geralmente 0) naqueles pontos.
- Se não existem graus de liberdade suficientes para impor condições adicionais, uma solução pode ser criar intervalos adicionais, para aumentar o número de graus de liberdade.
- Outra solução pode ser mudar a localização dos pontos extremos. Por exemplo, no caso da fecundidade, a curva não precisa começar necessariamente aos 15 anos, mas pode começar aos 14 ou 16 anos e terminar aos 49 ou 51 anos.
- Finalmente, é possível mudar o tipo da curva nos intervalos extremos. Por exemplo, no caso de uma interpolação linear da fecundidade, pode ser impossível encontrar uma especificação com valores não negativos tanto em $x=15$ como em $x=50$. Mas mudando a curva de linear para quadrática nos intervalos extremos cria graus de liberdade adicionais que podem permitir a obtenção de valores mais plausíveis nos extremos.

Como as quantidades demográficas geralmente são não negativas, as funções que resultam quando são acumuladas em intervalos sucessivos geralmente são crescentes. Outras funções, como o ℓ_x da tábua de vida necessariamente são decrescentes. Portanto, os seus interpolantes precisam refletir essa mesma tendência. O procedimento mais comum para garantir a monotonicidade envolve o uso de parâmetros livres nos nodos para poder escolher as derivadas nesses pontos. O mesmo procedimento também pode ser útil para regularizar outros aspectos do comportamento da curva interpolada. Essa possibilidade surge naturalmente quando se usam splines cúbicos de Hermite. Normalmente não é difícil aplicar a ideia de forma intuitiva, mas um método mais formal baseado nela foi sugerido por Hyman (1983) e introduzido na interpolação de quantidades demográficas por Smith, Hyndman e Wood (2004). O algoritmo aproveita o fato de que (18.23) e (18.24) permitem escolher os valores de $f'_i(x_i)$ e $f'_i(x_{i+1})$ de tal forma que o resultado seja monótono. O método, que eles chamam o “filtro de Hyman”, foi implementado em “R” e está disponível em:

<http://www-personal.buseco.monash.edu.au/~hyndman/Rlibrary/>.

Outro pacote desenvolvido pelo mesmo autor, chamado Smooth Demogdata, suaviza um conjunto de valores observados com splines monótonos, mas também oferece outras alternativas:

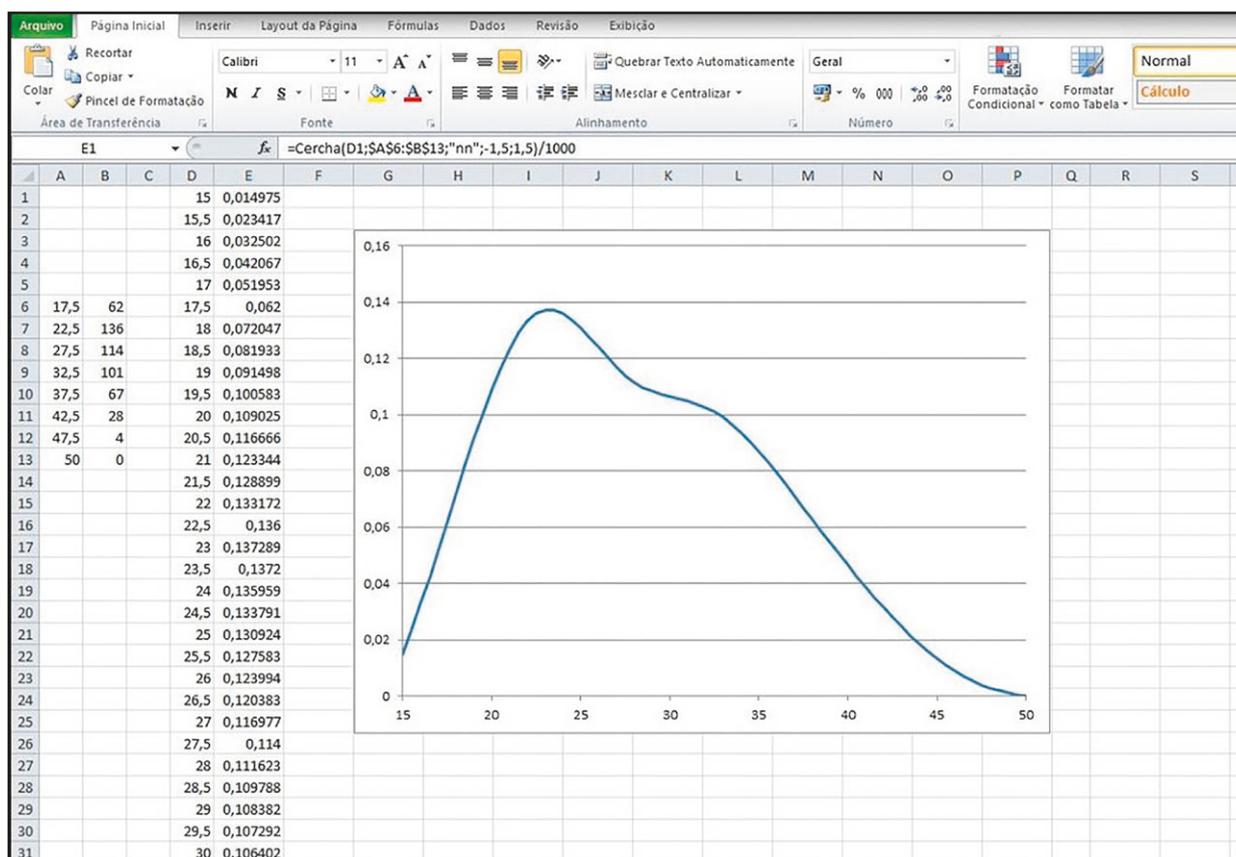
<https://www.rdocumentation.org/packages/demography/versions/1.20/topics/smooth.demogdata>.

O filtro de Hyman também está disponível na função `splinefun` da versão padrão de “R”, junto com o spline natural e algumas outras opções. Esta função não requer a instalação de bibliotecas adicionais.

18.5 SOFTWARE E RECURSOS PARA A INTERPOLAÇÃO

Até aqui todas as interpolações mostradas no texto foram geradas diretamente no EXCEL, sem qualquer recurso adicional. Entretanto, existem vários recursos, tanto no EXCEL como em “R”, que ajudam a fazer diferentes tipos de interpolação automaticamente. Um dos mais simples e acessíveis é o programa Interpolation.xla, de autoria de José Iván Martínez García, que está livremente disponível em internet (<https://personales.gestion.unican.es/martinji/Interpolation.htm>), num sítio com instruções em inglês e espanhol, e que pode ser baixado como Add-in de EXCEL. O programa contém várias funções, entre as quais, funções que fazem a interpolação com splines cúbicos naturais, splines de Hermite e splines de Akima (não discutidos aqui). O programa também contém várias funções utilitárias, por exemplo para diferenciar diferentes tipos de splines. A Figura 18.1 mostra uma imagem de tela para o cálculo do spline natural do Gráfico 18.5 com o uso da função Cercha de Interpolation.xla, onde os dados estão em A6:B13 e os valores interpolados na coluna E. O argumento “nn” da função Cercha se refere à opção de calcular um spline natural, com $f''(x) = 0$ em ambos os extremos.

Figura 18.1: Imagem de tela (ecrã) do spline cúbico natural do Gráfico 18.5 elaborado com a função Cercha de Interpolation.xla



Fonte: Censo Demográfico de Cabo Verde (2010).

Um “Add-in” semelhante que executa diferentes interpolações cúbicas, incluindo o spline de Bessel e o spline OneWay, e que também está disponível gratuitamente para ser baixado da internet, é SRS1 (<http://www.srs1software.com/SRS1CubicSplineForExcel.aspx>). O mesmo vale para um recurso um pouco mais elementar chamado XIXtrFun.xll que está disponível gratuitamente em <http://www.xlxtrfun.com/XIXtrFun/ReadMeXIXtrFunAndSurfGen.htm>). Embora esses recursos salvem tempo na implementação dos algoritmos rotineiros, eles têm a desvantagem de ser relativamente rígidos e de não permitir as variações dos métodos que às vezes se tornam necessárias para acomodar situações especiais, como condições específicas que devem ser satisfeitas nos extremos da curva.

Existem também alguns programas comerciais na área de interpolação e ajuste de dados. O programa Data Curve Fit Creator também tem o formato de um “Add-in” para EXCEL. Mais informação sobre este programa, que é comercializado pela mesma companhia responsável pelo SRS1, pode ser obtida no sítio de web <http://www.srs1software.com/DataCurveFitCreator.aspx>. O programa TableCurve 2D (ver em <https://systatsoftware.com/downloads/download-tablecurve-2d/>) tem finalidades mais gerais de ajuste de curvas, com um acervo de mais de 3.500 funções paramétricas e vários tipos de splines que podem ser ajustados aos dados observados.

Como sempre acontece com aplicativos de software que facilitam cálculos relativamente complexos, estes têm que ser usados com bastante critério. Quando usados com conhecimento de causa, eles podem ajudar a obter os resultados corretos com maior celeridade e flexibilidade. Mas quando usados de forma cega, sem conhecer os critérios aplicados, eles podem levar a resultados absurdos, sem que o usuário se dê conta disso.

