

## Módulo 7

# Estimação de parâmetros

*It is rarely possible to study the entire population, rarely possible to assess age intervals with certainty, and, for long-lived animals, rarely possible to study a cohort from birth to the death of its last member.*

*Dempster, J.R. 1975*

### 7.1 Estimação da fertilidade

Definimos fertilidade como o número de fêmeas-filhas viáveis produzidas por fêmea adulta por unidade de tempo (em geral o ano), a definição mais habitual em demografia. O termo fecundidade é em geral reservado para definir *potencial* reproductivo. O potencial é condicionado pela capacidade máxima de produzir gâmetas ou de produzir zigotos e é em geral mais fácil de medir que a fertilidade. Muitas vezes a fecundidade é proporcional ao tamanho corporal das fêmeas – uma associação que é bem conhecida, por exemplo, em insectos e em peixes. Uma vez que o tamanho corporal e a idade, por sua vez, também estão muito associados, o conhecimento da estrutura etária da população conduz, através destas relações, a uma estimativa da fecundidade global da população. Um certo número de animais na idade  $x$ , sendo  $x$  uma idade pós-primeira maturação, deve corresponder a uma certa fecundidade global da idade  $x$ .

Estimar a fertilidade é um pouco mais difícil por duas razões. Primeiro, porque é necessário saber qual a proporção de cada grupo etário que efectivamente se reproduz numa época de reprodução. Os biólogos das pescas, por exemplo, têm-se esforçado bastante para estimar a proporção da biomassa pós-idade de 1ª maturação que realmente desova em cada ano, um factor importante para conhecer a fecundidade que em cada ano se traduz em fertilidade.

Segundo, porque a fertilidade obriga a conhecer a proporção de ovos ou embriões formados que é “viável”. Não há à partida convenção sobre aquilo a que se chama fêmeas-filhas viáveis. Se são os zigotos femininos, ou se são as sobreviventes desde o zigoto até outro estado qualquer do ciclo de vida em que se contabilizam as filhas. Este outro estado pode ser o juvenil (individuos morfologicamente semelhantes a adultos mas que ainda não se reproduzem) ou pode até ser o estado de 1ª maturação. Seja como for, é necessário ter uma estimativa da sobrevivência no período que decorre entre o zigoto e o ponto do ciclo de vida em que se contabilizam as filhas. No caso das populações aquáticas com reprodução externa, é muito difícil estimar a sobrevivência dos ovos e das larvas, com vista a conhecer o volume de novos recrutas juvenis à população.

Em mamíferos, é muitas vezes mais fácil contabilizar o número de gravidezes por fêmeas com  $x$  anos de idade do que propriamente o número de filhos viáveis. Para se converter os numeros de

gravidezes em  $m_x$ , é necessário ter em atenção qual é a proporção de gravidezes que têm sucesso, o número médio de nados-vivos por gravidez e, finalmente, a taxa de sobrevivência dos recém-nascidos durante o intervalo de tempo que separa dois recenseamentos da população. No caso da espécie humana, é habitual encontrar a informação relevante para calcular  $m_x$  directamente nas publicações dos institutos estatais que coligem estatísticas da população (Figura 7.1).

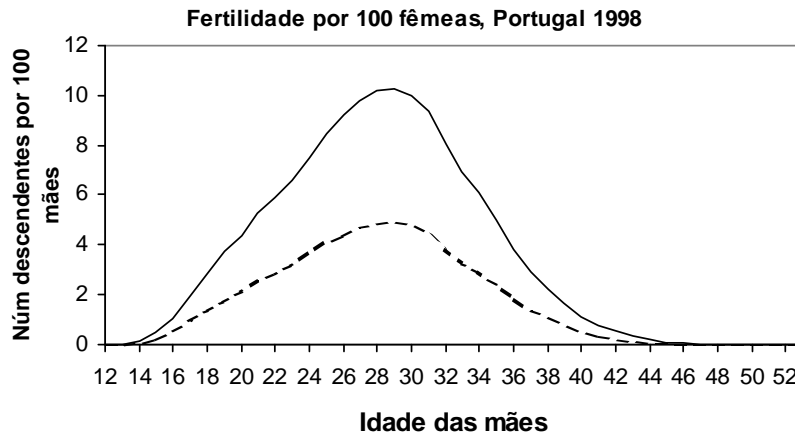


Figura 7.1 Fertilidade ( $m_x \times 100$ ) em Portugal em 1998. Número total de nados-vivos dos 2 sexos (a cheio) e femininos (a tracejado) por cada 100 mães. Em 1998 houve um total de 113500 nados-vivos. O seu sex-ratio foi 48.4% de fêmeas para 51.6% de machos. Fontes: Filhos por idade - de INE (1998); número de mães por idade - de INE (1997).

A fertilidade depende da idade e por vezes depende ainda mais do tamanho corporal, especialmente em animais ectotérmicos. Os juvenis, por definição, não se reproduzem e, em quase todas as espécies, a duração deste estágio pré-reprodutor não pode ser negligenciada quando se projecta a população. A taxa instantânea de crescimento da população ( $r$ ) é bastante sensível à idade de 1ª reprodução – o seu adiamento diminui a taxa de crescimento de forma significativa. Assim sendo, a utilização de um único número para representar a taxa de natalidade ( $b$ ) é uma sobre-simplificação que ignora o período juvenil e pode conduzir a uma sobre-estimação grosseira da taxa instantânea de crescimento da população. Não existe um bom substituto para o vector com as fertilidades de todas as idades,  $[m_x]$ . Caso seja necessário usar um único número que represente a taxa de natalidade,  $b$ , provavelmente o melhor é usar o vector  $[m_x]$  completo para estimar  $r$  e depois subtrair a  $r$  a mortalidade global ( $d$ ), a fim de obter  $b$ .

A estimação da fertilidade em plantas não é necessariamente menos complicada do que nos animais. Muitas vezes, envolve a contagem das sementes produzidas por um indivíduo por unidade de tempo. Para determinar o número de sementes, o habitual não é contá-las todas mas sim tomar uma amostra de indivíduos e uma sub-amostra de ramos, frutos, ou o que fôr apropriado, dentro de cada indivíduo. O número médio de sementes por ramo (ou fruto) é calculado, depois é multiplicado pelo número médio de ramos (ou frutos) por indivíduo e, finalmente, pelo número de indivíduos. Em princípio, isto seria feito para indivíduos da mesma idade, mas é frequente trabalhar-se com classes de tamanho em vez de idade. Para se chegar a  $m_x$  é necessário ainda estimar qual é a proporção de sementes que germina e qual é a sobrevivência das que germinam. Nada simples. As coisas complicam-se um pouco mais quando no mesmo ciclo de vida também existe reprodução assexuada.

Note-se que a definição de  $m_x$  continua a ser válida para a reprodução assexuada. O valor de  $m_x$  total é a soma dos descendentes sexuados e assexuados por fêmea de idade  $x$ .

Apesar das dificuldades, em geral, a fertilidade é um dos parâmetros demográficos para o qual é mais fácil dispor de informação que conduza a uma estimativa operacional, pelo menos em aves e mamíferos. Podem-se utilizar conhecimentos provenientes de animais em cativeiro sobre a duração da sua gestação e a fertilidade para aplicar a populações selvagens, sem que o erro cometido seja muito grande. Mas a estimação da função  $m_x$  em populações animais tem um alto grau de especificidade, dependendo muito da espécie em causa. O leitor interessado encontra mais detalhes sobre a quantificação da fertilidade em Southwood (1978), Carey (1993), Ebert (1999), Skalski et al (2005).

## 7.2 Estimação do tempo de geração, T.

Recorde-se que T é o tempo médio decorrido entre o nascimento dum indivíduo e o nascimento dos seus filhos. Podemos calculá-lo de forma aproximada, calculando a idade média da coorte no momento em que os indivíduos têm filhos. Sabemos já que um indivíduo tem, em média,  $l_0m_0$  filhos na idade  $x=0$ , tem  $l_1m_1$  filhos na idade  $x=1$ , tem  $l_2m_2$  na idade  $x=2$ , etc. A soma das idades da coorte nos momentos de ter filhos é, portanto, o somatório dos  $l_xm_x x$  para todas as idades. Dividindo este somatório pelo número médio total de filhos tidos, tem-se,

$$T \approx \frac{\sum_{x=0}^L l_x m_x x}{\sum_{x=0}^L l_x m_x} = \frac{\sum_{x=0}^L l_x m_x x}{R_0} \quad [7.1]$$

sendo L a idade mais avançada com indivíduos vivos na LT e  $\approx$  lê-se “aproximadamente igual a”.

A equação [7.1] é uma média ponderada. Deve ser vista como a média de  $x$ , ponderada pela frequência  $l_x m_x$ . Porque razão é uma aproximação? Embora na LT a idade seja apresentada em saltos discretos, na verdade  $x$  é uma variável contínua, o mesmo se passando com  $l_x$  e  $m_x$ . Em termos matemáticos, deve-se pensar na distribuição da densidade de  $l_x m_x$  com respeito a  $x$ . O cálculo rigoroso de T é feito substituindo os somatórios em [7.1] por integrais em ordem à idade,

$$T = \frac{\int_0^L l_x m_x x dx}{\int_0^L l_x m_x dx} \quad [7.2]$$

Evidentemente, na prática não se pode usar [7.2] porque a função  $l_x m_x$  não é dada pela LT e não é conhecida, logo não se pode integrar.

A equação [7.1] ignora a estrutura etária da população. Imagine-se, por exemplo, que as idades  $x=0$  e  $x=1$  compreendem cerca de 90% da população e que as idades  $x=2$  e  $3$ , embora tendo indivíduos presentes, estes representam os restantes 10% da população. Faz sentido dar o mesmo peso a todas as idades no cálculo de  $T$ ? Alguns autores ponderam cada parcela dos somatórios em [7.1] pela abundância da respectiva idade em DEE. Recordando que em DEE a expressão  $e^{-rx}$  determina a abundância relativa de cada grupo de idade na população, o tempo médio de geração é estimado por,

$$T_{DEE} \approx \frac{\sum_{x=0}^L e^{-rx} l_x m_x x}{\sum_{x=0}^L e^{-rx} l_x m_x} \quad [7.3]$$

Note-se que se a população estiver estacionária (ZPG),  $r=0$  e  $e^{-rx}=1$ , fazendo com que as equações [7.1] e [7.3] dêem o mesmo resultado. As diferenças entre  $T$  e  $T_{DEE}$  em geral são pequenas, mas tendem a aumentar quando a população tem uma taxa de crescimento elevada ( $r \gg 0$ ). Por exemplo, em populações com  $r > 1$ , a pirâmide etária tende a ser dominada pelas idades mais jovens e, nesse caso,  $T > T_{DEE}$ . Em populações humanas, a média aritmética das duas formas de calcular o tempo de geração é um bom estimador do verdadeiro tempo de geração:  $(T+T_{DEE})/2$  (Keifitz and Caswell 2005).

Há ainda uma terceira forma de calcular  $T$  que combina as taxas vitais da coorte com o crescimento observado da população. Se se definir  $T$  como sendo o tempo que a população leva para crescer a quantidade  $R_0$ . O crescimento da população ao fim de  $n$  unidades de tempo é dado por:

$$N_{t+n} = N_t \lambda^n$$

quando  $n$  corresponde a  $T$ , o tempo de 1 geração, então  $\lambda^T$  corresponde a  $R_0$ , ou seja,

$$N_{t+T} = N_t \lambda^T = N_t R_0$$

da relação  $\lambda^T = R_0$  retira-se:

$$T = \frac{\ln R_0}{\ln \lambda} \quad [7.4]$$

### 7.3 Estimação de $r$

O parâmetro  $r$  pode ser interpretado de duas formas (McCallum 2000). Uma é a taxa intrínseca de crescimento máximo que uma população pode ter, quando não existem limitações de recursos alimentares ou espaciais. Neste caso  $r=b_0-d_0$ , onde  $b_0$  e  $d_0$  são, respectivamente as taxas de natalidade máxima e de mortalidade mínima. É este o valor que se espera que a população tenha em condições ambientais ótimas, quando não existem constrangimentos de regulação dependente ou independente da densidade. Por exemplo, quando a população coloniza o habitat e está a crescer a partir dum nível

muito baixo. Esta quantidade poder-se-ia designar por  $r_0$  ou  $r_{máx}$  e é uma propriedade, simultaneamente, da espécie e do habitat em que esta está.

Outra forma de interpretar  $r$  é como a taxa de crescimento adoptada pela população para um dado conjunto de parâmetros da Life Table, nomeadamente os vectores  $[l_x]$  e  $[m_x]$ , logo que a população atinge a Distribuição Etária Estável (DEE). Este valor, que designarei por  $r_s$ , em geral é inferior a  $r_{máx}$ , por definição de  $r_{máx}$ . Pode haver factores de regulação da densidade populacional a actuar que tornam  $r_s$  muito inferior a  $r_{máx}$ , mas raramente é possível saber a que distância  $r_s$  está de  $r_{máx}$ . A maior parte dos autores, infelizmente, apenas usa  $r$ , sem distinguir entre  $r_s$  e  $r_{máx}$ , sendo a partir do contexto que se compreende a qual das duas taxas se referem.

No módulo 6, mostrou-se que uma população com taxas vitais constantes ao fim de alguns anos entra em DEE e, a partir desse momento, cresce geometricamente. A taxa instantânea de crescimento,  $r$ , pode ser estimada a partir de duas observações consecutivas do efectivo duma população nessas circunstâncias. Calcula-se primeiro a taxa finita de incremento  $\lambda = N_{t+1}/N_t$  e, depois,  $r = \ln \lambda$ . Na ausência de observações do efectivo populacional em DEE, porém, ainda se pode estimar  $r$  a partir das taxas vitais presentes na Life Table (LT). Este  $r$  será  $r_s$  ou  $r_{máx}$  dependendo do contexto em que a população se encontra, respectivamente, sujeita a restrições de crescimento ou sem quaisquer restrições. Na esmagadora maioria dos casos estar-se-á a estimar  $r_s$ .

Há dois métodos principais para estimar  $r$  a partir da LT, nenhum deles requerendo que os intervalos  $\Delta x$  na LT sejam todos iguais. Um, mais simples e aproximativo, utiliza  $R_0$ , a taxa líquida de reprodução. O outro é rigoroso, mas biologicamente menos transparente e computacionalmente mais exigente. Começo pelo primeiro.

Como vimos,  $R_0$  é o factor multiplicativo que permite calcular o efectivo populacional entre duas gerações consecutivas. Se a população cresce geometricamente, a relação entre o número de indivíduos no instante  $t$  e o número de indivíduos  $T$  tempo mais tarde, sendo  $T$  o tempo que dura uma geração, é dada pela equação (recordar o crescimento exponencial):

$$N_{t+T} = N_t e^{rT} \quad [7.5]$$

Uma vez que  $R_0 = N_{t+T}/N_t$ , obtem-se a relação entre  $r$  e  $R_0$ ,

$$R_0 = e^{rT} \quad \text{ou} \quad r = \frac{\ln R_0}{T} \quad [7.6]$$

Os valores de  $R_0$  e de  $T$  são estimados a partir da LT e do crescimento populacional, conforme já descrito.

O segundo método para calcular  $r$  tira partido da equação [6.2] (módulo 6):

$$B_t = B_0 e^{rt} \quad [7.7]$$

que relaciona  $B_t$ , o número de nascimentos no corrente ano  $t$ , com o número de nascimentos  $t$  anos atrás,  $B_0$ . O número de indivíduos na idade  $x$  do ano  $t$  é então  $B_0 l_x$ , ou seja, o número de nascimentos  $x$  anos atrás que sobrevive até  $t$ . Substituindo  $B_0$  pela sua expressão deduzida a partir de [7.5], e fazendo  $t$  corresponder a  $x$  anos,

$$B_t e^{-rx} l_x \quad [7.8]$$

(Os estudantes com dúvidas devem rever o Módulo sobre Estrutura Etária). No ano  $t$ , cada um destes sobreviventes contribui para os nascimentos com  $m_x$  descendentes. O número total de contribuições da idade  $x$  no ano  $t$  é portanto  $m_x l_x B_t e^{-rx}$ . O número total de nascimentos,  $B_t$ , é a soma da contribuição de todas as idades,

$$B_t = m_0 l_0 B_t e^{-r0} + m_1 l_1 B_t e^{-r1} + \dots + m_L l_L B_t e^{-rL} = B_t \sum_{x=0}^L m_x l_x e^{-rx} \quad [7.9]$$

Dividindo tudo por  $B_t$ ,

$$1 = \sum_{x=0}^L m_x l_x e^{-rx} \quad [7.10]$$

Isto é a chamada **equação de Lotka**, designada também por alguns autores equação de Euler, uma vez que foi deduzida por Leonhard Euler (1707-1783) e Alfred Lotka (1880-1949). Conhecidos  $l_x$  e  $m_x$ , a única incógnita em [7.10] é  $r$ . Infelizmente,  $r$  não pode ser isolado num dos lados da equação, portanto não pode ser calculado de imediato. O valor de  $r$  tem de ser calculado por iteração de [7.10], como exemplifico adiante. À parte esse inconveniente, o valor de  $r$  obtido é o mais correcto.

A equação [7.6] é uma relação de fácil interpretação biológica entre a taxa intrínseca de crescimento,  $r$ , e outros parâmetros,  $R_0$  e  $T$ , pelo que a equação tem sido um valioso ponto de partida em discussões acerca do significado evolutivo das estratégias de vida (Pianka 1978, Roughgarden 1979, Yodzis 1989). A equação de Lotka, pelo contrário, embora apareça em todos os tratados de ecologia, não é muito adequada para uma discussão biológica das relações entre  $r$  e outros parâmetros sinópticos do ciclo de vida da espécie. A relação que existe entre [7.6] e a equação de Lotka é desde há muito conhecida em demografia matemática (e.g. Pollard 1973, May 1976). A qualidade da aproximação que [7.6] representa relativamente ao verdadeiro  $r$ , depende da forma da função  $f(x)=l_x m_x$ . Demonstra-se que [7.6] constitui uma boa aproximação quando  $R_0 \approx 1$  (a densidade populacional permanece aproximadamente constante, como é o caso de muitas populações humanas), quando a variância da função  $l_x m_x$  é pequena (i.e. os nascimentos de filhos estão muito concentrados em torno duma idade da mãe), ou quando há uma combinação das duas coisas. Quanto à equação de Lotka, a sua qualidade como estimador de  $r$  requer apenas que  $f(x)$  não tenha assimetrias que a desviem demasiado da distribuição Normal. Mas mesmo com assimetrias, conduz a um  $r$  que, em geral, é preciso até 3 ou mais casas decimais. Os detalhes matemáticos da demonstração do que acabo de dizer extravasam o âmbito deste curso.

Para finalizar esta secção, vamos calcular  $r$  usando os dois métodos. A Tabela 7.1 é a LT duma espécie com 3 idades, na qual foram acrescentadas as colunas,  $l_x m_x$  e  $l_x m_x x$ , que são necessárias aos cálculos.

Tabela 7.1 Life Table com as colunas necessárias para o cálculo de  $r$

Idade ( $x$ )	$l_x$	$S_x$	$m_x$	$l_x m_x$	$x l_x m_x$
0	1.000	0.240	0	0.000	0.000
1	0.240	0.242	20	4.800	4.800
2	0.058	0.000	24	1.392	2.784
3	0.000			0.000	0.000
Soma:				6.192	7.584

A partir da tabela, calcula-se directamente:

$$R_0 = \sum_{x=0}^L l_x m_x = 6.192 ; \quad T = \frac{\sum x l_x m_x}{R_0} = \frac{7.584}{6.192} = 1.225 \text{ anos}$$

donde,

$$r = \frac{\text{Ln } R_0}{T} = \frac{\text{Ln } 6.192}{1.225} = 1.489 \text{ ano}^{-1}$$

Este valor de  $r$  é diferente do valor  $r=1.624$ , obtido no módulo 6 quando se projectou, ano a ano, uma população com esta LT. Vejamos agora a utilização da equação de Lotka, recorrendo a um método iterativo muito simples. O método consiste em começar por substituir no lado direito de [7.10] um valor aproximativo de  $r$ , o qual pode ser, por exemplo, o que acabámos de estimar:  $r=1.489$ . Em seguida, efectuam-se as contas do lado direito de [7.10] e compara-se com o lado esquerdo da equação. O valor obtido é 1.154, o qual é superior a 1. A Tab. 7.2 ilustra as contas.

É necessário portanto tentar um valor de  $r$  novo, capaz de fazer 1.154 ficar mais próximo de 1. Para conseguir que o lado direito de [7.10] dê um valor mais baixo, deve-se introduzir um  $r$  mais *alto*, uma vez que  $r$  está num expoente de sinal negativo. Nesta segunda tentativa, tentei  $r=1.6$ . O resultado do lado direito de [7.10] foi 1.026, que ainda é superior a 1. Repeti então com  $r=1.65$ , que deu 0.973 (<1) e, portanto, introduzi a seguir um  $r$  mais *baixo*. A Tabela 7.3. resume os resultados de várias tentativas feitas até se conseguir a estimativa  $r=1.624$ , que coincide com o verdadeiro valor de  $r$ .

Tabela 7.2 Cálculos do lado direito da equação [7.8], com o valor tentativo inicial de  $r=1.489$ . O resultado dos cálculos é 1.154.

Idade ( $x$ )	$l_x m_x$	$\exp(-1.489 x)$	$l_x m_x \exp(-1.489 x)$
0	0.000	1.000	0.000
1	4.800	0.226	1.083
2	1.392	0.051	0.071
3	0.000	0.011	0.000
Soma:			1.154

Tabela 7.3 Iterações sucessivas com a equação de Lotka, seguindo a regra: se a coluna mais à direita dá um valor superior a 1, o valor de  $r$  seguinte deve aumentar e vice-versa.

Número da iteracção	$r$	$l_x m_x \exp(-r x)$
1	1.489	1.154
2	1.600	1.026
3	1.650	0.973
4	1.630	0.994
5	1.625	0.999
6	1.624	1.000

Normalmente, esta forma de lidar com a equação de Lotka não coloca dificuldades práticas, dada a actual disponibilidade de meios computacionais. Convém contudo esclarecer que há formas muito mais eficientes de lidar com equações em que a incógnita não se pode explicitar. Para os mais interessados, o assunto é tratado num ramo da matemática chamado análise numérica.

*É assim tão importante estimar  $r$  ?*

McCallum (2000) argumenta que não. Os modelos biomatemáticos que usam  $r$  têm um razoável grau de falta de detalhe ou, se se preferir, de abstracção. É o caso da equação do crescimento exponencial,  $N_t = N_0 e^{rt}$ , ou da equação logística de Verhulst. Estes modelos pretendem apenas descrever de forma aproximativa o crescimento populacional e talvez não faça muito sentido investir muito tempo ou dinheiro na determinação de  $r$  até mais do que 1 casa decimal para usar em modelos destes. Quase sempre, a projecção da população é feita de forma mais adequada subdividindo-a em estádios (ou idades) e projectando com as sobrevivências e fertilidades específicas de cada estádio (ou idade). Este tipo de conhecimento, por sua vez, tem como *output* o próprio  $r$ . Evidentemente, se não for possível trabalhar com a população estruturada em estádios, então estimar  $r$  poderá ser mesmo importante.



## 7.4 Estimação da sobrevivência

Nas populações selvagens, a sobrevivência é em geral mais difícil de quantificar que a fertilidade, uma vez que raramente a informação proveniente de indivíduos em cativeiro serve de indicador para a sobrevivência fora de cativeiro. Poderá, contudo, dar uma indicação da longevidade máxima que os indivíduos podem atingir. Idealmente, estimar a sobrevivência no campo devia ser feito acompanhando as coortes ao longo da vida. Na prática isto é difícil ou mesmo impossível e pode-se, em alternativa, estimar a sobrevivência por idade a partir da estrutura etária da população sob pressupostos que já são familiares. Admitindo, evidentemente, que é possível conhecer a idade dos indivíduos o que, por si só, pode ser um assunto tecnicamente difícil.

Vamos agora examinar técnicas de estimação das taxas de sobrevivência. Quando apresentei a Life Table (LT), assumi que os dados de base necessários à sua construção, nomeadamente as taxas de sobrevivência por idade, eram informação conhecida do biólogo e isenta de problemas. Infelizmente, a realidade não é assim. Quando pretendemos fazer a análise demográfica de uma população, é quase certo que teremos de investir mais tempo, dinheiro e paciência na estimação das taxas de sobrevivência do que propriamente na análise da LT. Uma vez isto deve-se a que a população é demasiado grande ou tem demasiadas idades, outras vezes os indivíduos são raros ou difíceis de seguir no seu habitat natural, outras ainda é porque há dificuldade em determinar a sua idade. Frequentemente, o biólogo tem de recorrer a métodos imaginativos para conseguir construir a sua LT e, uma vez que estes dependem grandemente da espécie em causa, do tempo e do dinheiro disponível, não pretendo falar de forma exaustiva sobre este assunto.

Antes de prosseguir, aqui fica um breve formulário do zoo de equações que o leitor já deve ter encontrado (módulo 2), relativas ao cálculo a taxas de sobrevivência e mortalidade. Para taxas de sobrevivência e mortalidade totais:

$$S_t = \frac{N_t - D_t}{N_t} = \frac{N_{t+1} - B_t}{N_t} \quad [2.4]$$

$$S_t = \frac{N_t + B_t - D_t}{N_t + B_t} = \frac{N_{t+1}}{N_t + B_t} \quad [2.5]$$

$$\frac{N_{t+\Delta t} - B_t}{N_t} = S_t = e^{-d\Delta t} \quad [2.20]$$

$$S_t = e^{-d} \quad \text{ou} \quad d = -\ln S_t \quad [2.21]$$

Para taxas de sobrevivência e mortalidade por idade (módulo 4):

$$S_x = \frac{N_{x+1}}{N_x} \quad [4.1]$$

$$l_x = \frac{N_x}{N_0} \quad [4.2]$$

$$q_x = 1 - S_x = \frac{N_x - N_{x+1}}{N_x} = \frac{D_x}{N_x} \quad [4.5]$$

$$N_{x+\Delta x} = N_x e^{-\mu \Delta x} \quad [4.9]$$

$$l_x = \frac{N_x}{N_0} = e^{-\mu x} \quad [4.11]$$

$$S_x = e^{-\mu x} \quad i.e. \quad \mu_x = -LnS_x \quad [4.14]$$

$$N_{xm} = \frac{N_x}{\mu_x} (1 - e^{-\mu_x}) \quad [4.15]$$

### 7.5 A curva de sobrevivência por idade: quatro métodos

Começo por distinguir quatro formas principais de reunir informação necessária ao cálculo das taxas de sobrevivência *por idade*, necessárias à construção da LT (eqs. [4.1] e [4.2]) ou da curva de sobrevivência (Mód. 4, Sec. 4.3),

- (1) Acompanhamento (longitudinal) de uma coorte,
- (2) Estimação a partir da estrutura etária (vertical) no ano  $t$ ,
- (3) Estimação a partir de duas (ou mais) observações da estrutura etária da população em  $t$ ,  $t+\Delta t$ , ... etc,
- (4) Acompanhamento de coortes diferentes em intervalos de tempo diferentes (Kaplan Meier e Nelson-Aalan).

O método (1) é o mais simples e natural. Começa-se com um conjunto de indivíduos de idade 0 e segue-se a vida desta coorte até ao último sobrevivente, calculando a sobrevivência de idade para idade. Este método permite a construção da LT horizontal.

O método (2) consiste em estimar a estrutura etária da população num instante de tempo (na prática, um curto intervalo de tempo) e depois, assumindo que a população está estacionária, estima-se a sobrevivência como no método (1). Este método conduz à LT vertical.

O método (3), intermédio entre os dois anteriores e está resumido na Figura 5.2 (módulo 5). Consiste em estimar a estrutura etária no instante de tempo  $t$ :  $N_{x,t}$  ( $x=0, 1, \dots$ ), voltar a estimá-la um intervalo de tempo mais tarde  $t+\Delta t$  (em geral  $\Delta t=1$  ano), obtendo  $N_{x,t+\Delta t}$  ( $x=1, 2, \dots$ ) e, eventualmente, fazer novas estimativas da estrutura etária em  $t+2, t+3, \dots$ . As sobrevivências por idade durante 1 ano, são então calculadas usando os pares sucessivos de valores  $N_{x+1,t+\Delta t}$  e  $N_{x,t}$ . O procedimento é feito para a transição entre os anos  $t$  e  $t+1$ ,  $t+1$  e  $t+2$ , etc., obtendo-se várias estimativas de  $S_x$ . Por exemplo, se se conhecer a população em 1 de Janeiro de 2006 e em 1 de Janeiro de 2007, a taxa de sobrevivência da idade  $x$  é simplesmente,  $S_x = N_{x+1,2007}/N_{x,2006}$ . Podia-se obter nova estimativa de  $S_x$  entre 2007 e 2008, caso se venha a conhecer a população em 2008.

O procedimento (4), é uma combinação dos métodos (1) e (3), muito usada na prática, dada a dificuldade em seguir coortes e em estimar a estrutura etária num ano com precisão aceitável. Há muitos estudos em que os animais entram em alturas diferentes e saem em alturas diferentes do estudo. Uma vez porque migram, outras porque se acabam as baterias dos rádio-emissores que lhes são colocados, ou ainda por qualquer razão que impeça que sejam seguidos, dão origem àquilo que em estatística se designa por dados “censurados” ou “truncados”<sup>1</sup>. Vou considerar duas variantes para lidar com este tipo de dados – o método Kaplan-Meier e o método Nelson-Aalan. Para melhor compreensão, apresento-os em seguida, no contexto de um exemplo de aplicação em ursos cinzentos.

#### Método Kaplan-Meier

Raramente é possível acompanhar uma coorte de animais móveis, especialmente se são muitos e vivem muitos anos. Por vezes, contudo, é possível acompanhar animais de coortes diferentes, quer simultaneamente, quer espaçadamente ao longo do tempo. Se se assumir que os factores ambientais que afectam a mortalidade actuaram sobre todas as coortes (parcialmente) acompanhadas de forma aproximadamente idêntica, é possível combinar toda a informação disponível e reconstruir a curva de sobrevivência da população em causa. O exemplo seguinte, com a população de ursos cinzentos dum estado americano, ilustra a situação e os procedimentos a adoptar.



O urso cinzento (*Ursus arctos horribilis*) é um dos ursos mais agressivos que existe. Não tem predadores nem inimigos naturais. Estes animais são omnívoros e passam a Primavera e Verão em busca de alimento tão variado como bolotas, amoras, mel, insectos, peixe, roedores e outros pequenos mamíferos. No Outono preparam um local de abrigo onde, em geral, passarão grande parte do Inverno a hibernar. Se, contudo, o Inverno não for muito rigoroso, fazem saídas ocasionais em busca de alimento. A longevidade destes animais ronda os 30 anos.

<sup>1</sup> A diferença entre estes dois termos é quase semântica. Os dados dizem-se *censurados* quando a vida de um indivíduo só é conhecida num certo intervalo de tempo. Dizem-se *censurados à direita* se tudo o que se sabe do indivíduo, a partir de certa altura, é que ainda está vivo, *censurados à esquerda*, se o indivíduo já tinha a característica em estudo (e.g. estar vivo) antes do estudo se iniciar. Há *truncagem à esquerda*, se alguns indivíduos vivos não foram deliberadamente incluídos no início do estudo. Há *truncagem à direita* se só os indivíduos que tiveram a característica de interesse (e.g. morrer) até um tempo pré-especificado, são incluídos na amostra (Klein and Moeschberger 1997).

O “Department of Fish and Game” (DFG) do estado americano do Idaho levou a cabo um estudo dos ursos que incluiu a estimação das suas taxas de sobrevivência (Skalski *et al* 2005). Foram capturados, anestesiados, marcados com rádio-emissores e libertados, 28 ursos ao longo de onze anos (1983-1993). Os animais tinham entre 1.5 e 28.5 anos de idade e entraram no estudo em alturas diferentes. Quando um emissor deixa de emitir, por exemplo por se esgotar a bateria, regista-se o que aconteceu ao animal: ou morreu devido a causas naturais, ou morreu devido a acções humanas (e.g. caça furtiva), ou continua vivo. Neste último caso, diz-se que o animal foi “censurado à direita”, uma vez que deixa de poder ser acompanhado embora esteja vivo. A Fig. 1 resume o resultado do acompanhamento dos 28 ursos. Quando um urso é censurado à direita, não é contabilizado nas mortes. Por exemplo, os ursos 3 e 4 foram seguidos pela última vez quando tinham 11,5 anos de idade, altura em que as baterias dos seus rádio-emissores se esgotaram. Na Fig. 1, a sua barra horizontal é estendida até aos 12,5 anos, para indicar que, em princípio, não morreram aos 11,5 anos de idade.

A Fig. 7.2 mostra que os dois animais mais jovens, marcados com 1.5 anos de idade (ursos 7 e 19) não sobreviveram nem 1 ano. Em todas as outras idades com animais, pelo menos um sobreviveu mais de 1 ano. A utilização daqueles dois animais (n=2, uma amostra muito pequena) conduziria à estimativa  $S_{1,5}=0$ , pelo que os animais foram descartados e só se conta com os animais marcados a partir dos 2,5 anos. Um estudo que ignora animais “à esquerda” diz-se truncado à esquerda. Além disso há, como se viu, ursos que saem do estudo antes de morrerem, pelo que o estudo é também censurado à direita.

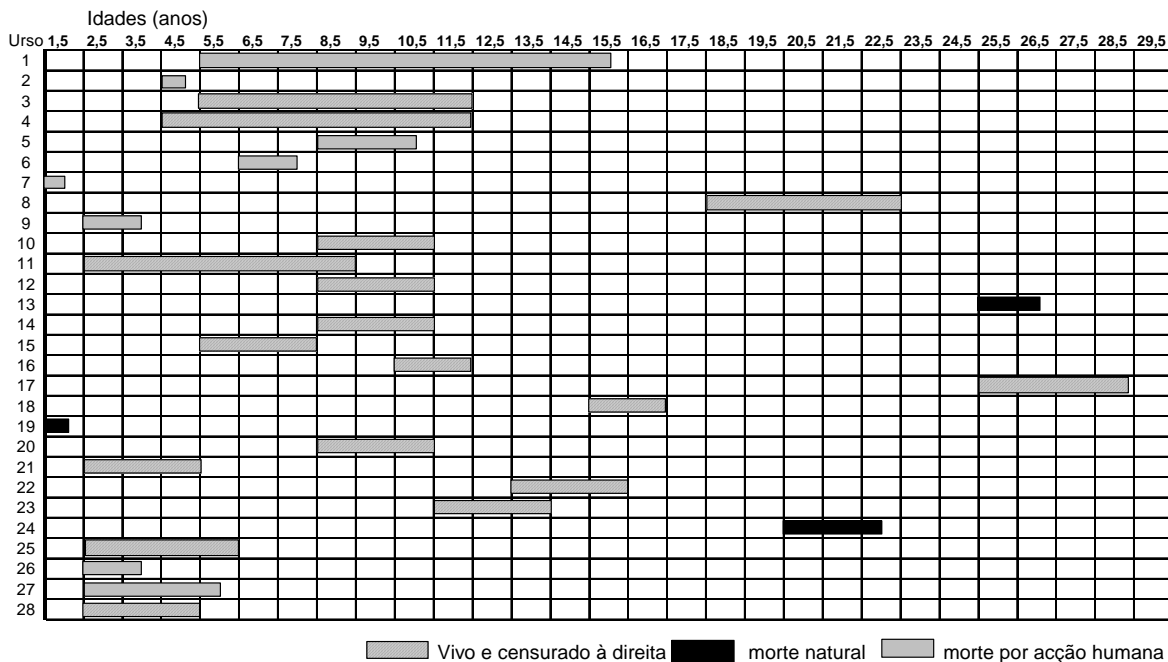


Figura 7.2. Idades de entrada e saída de 28 animais, num estudo de marcação com rádio-emissores que se prolongou por 11 anos. Quando um emissor parava, registava-se o destino do animal. A maioria continuava viva, mas não podiam continuar a ser acompanhados, pelo que se dizem “truncados à direita”. Os cálculos da sobrevivência ignoraram os dois animais com 1.5 anos (ursos 7 e 19) que entraram no estudo, pois estes morreram pouco depois, por isso o estudo diz-se também “truncado à esquerda”. Notar que o eixo de abcissas representa idade dos animais e não tempo do estudo. Por exemplo, o urso 17 está colocado na ponta direita, mas pode ter sido acompanhado primeiro que o urso 7 colocado à esquerda. Notar também que a idade dos animais está contabilizada em termos de “anos-e-meio”, uma vez que a época de marcações se localiza entre duas épocas de reprodução.

O método Kaplan-Meier (KM) é muito usado em epidemiologia e permite lidar com dados censurados. O estimador KM é, até certo ponto, já nosso conhecido,

$$l_x = \prod_{i_j < x} S_i \quad [7.11]$$

Onde  $i_1, i_2, \dots, i_k$  são pontos no tempo anteriores a  $x$ . Estes pontos não têm de ser igualmente espaçados no tempo, mas por razões simplificadoras vou assumir que são.

Pode-se demonstrar (Pollock *et al* 1989) que [7.11] não é só aplicável a situações em que os indivíduos “em risco” decrescem continuamente devido à mortalidade, como quando se acompanha uma coorte. A sobrevivência entre as idades  $x$  e  $x+1$ , é estimada com base nos animais que estão vivos na idade  $x$ , e estes podem aumentar ou diminuir. Por exemplo, retomem-se os ursos e considere-se a probabilidade de sobreviver entre as idades 10.5 e 11.5. Foram seguidos  $N_{10,5}=9$  animais (não necessariamente todos no mesmo ano!), dos quais sobreviveram 8, conduzindo à estimativa  $S_{10,5}=8/9 = 0.889$ .

A função de sobrevivência até à idade,  $l_x$ , pode ser construída a partir do produto das sucessivas estimativas de  $S_i$  (para  $i < x$ ). Por exemplo, a probabilidade de um urso vivo aos 2.5 anos chegar vivo aos 5.5 anos é,

$$l_{[2.5, 5.5]} = S_{2.5} S_{3.5} S_{4.5} = \left(\frac{7}{7}\right) \left(\frac{5}{7}\right) \left(\frac{6}{7}\right) = 0.612$$

A Tabela 1 resume os cálculos para obtenção de toda a curva de sobrevivência dos ursos, entre as idades 2.5 e 28.5, multiplicando gradualmente os valores de  $S_i$ . Quando, para determinada idade, não há animais ( $N_i = 0$ ), temos um problema. É o que acontece nas idades 17.5, 23.5 e 24.5. Nestes casos, os valores de  $S_i$  têm de ser presumidos, eventualmente com base nas sobrevivências imediatamente adjacentes. Evidentemente, é conveniente não avançar para a estimação de  $l_x$  se houver muitas idades nestas circunstâncias.

Notar, comparando a Fig. 7.2 e a Tab 1, que os ursos censurados à direita não são contabilizados como mortes. Por exemplo, no início dos 11,5 anos havia 5 ursos vivos e, embora 3 tenham deixado de ser seguidos nessa idade,  $S_{11,5}=5/5 = 1$ .

Tabela 1. Taxas de sobrevivência por idade ( $S_i$ ) e até à idade ( $l_x$ ) dos ursos cinzentos. Ignorou-se a idade 1.5 e seguiram-se os ursos a partir de  $i = 2.5$ , idade em que, por definição,  $l_x = 1$ . Em idades para as quais não existem dados (ND), foi necessário assumir um valor para  $S_i$ . Assumiu-se  $S_{17.5} = 1$ ,  $S_{23.5} = S_{24.5} = 0.75$ .

		Idades (anos)													
		1,5	2,5	3,5	4,5	5,5	6,5	7,5	8,5	9,5	10,5	11,5	12,5	13,5	14,5
$N_i$		2	7	7	7	7	6	6	9	8	9	5	2	3	2
$D_i$		2	0	2	1	1	0	1	0	0	1	0	0	0	0
$S_i$		0	1	0,714	0,857	0,857	1,000	0,833	1,000	1,000	0,889	1,000	1,000	1,000	1,000
$l_x$			1	1	0,714	0,6119	0,5244	0,524	0,437	0,4368	0,4368	0,3883	0,388	0,3883	0,3883

		15,5	16,5	17,5	18,5	19,5	20,5	21,5	22,5	23,5	24,5	25,5	26,5	27,5	28,5
$N_i$		3	1	ND	1	1	2	2	2	ND	ND	2	2	1	1
$D_i$		1	0	ND	0	0	0	0	1	ND	ND	0	1	0	0
$S_i$		0,667	1	ND(1)	1	1	1	1	0,5	ND(0,75)	ND(0,75)	1	0,5	1	1
$l_x$		0,3883	0,259	0,259	0,259	0,259	0,259	0,259	0,259	0,1295	0,0971	0,0728	0,073	0,0364	0,0364

### Método Nelson-Aalen

O método Kaplan-Meier pode ser um mau estimador da função  $l_x$  quando as amostras de que se dispõe são pequenas. Nos ursos, não pudemos começar na idade 1.5, devido a um relativo excesso de mortes ( $S_{2.5} = 0$ ), o qual anularia qualquer cálculo de  $l_x$  em que a idade 1.5 interviesse. Além disso, os estimadores de variância só estão definidos para idades com  $N_x > 0$ , devido à forma como  $N_x$  aparece nos denominadores. Mesmo quando  $N_x > 0$  para todas as idades, o carácter multiplicativo do estimador de  $l_x$ , a partir de valores de  $S_x$ , faz com que erros em estimativas de  $S_x$  se propaguem às outras estimativas.

Nelson (1972) e Aalen (1978) desenvolveram métodos não-paramétricos<sup>2</sup> alternativos para estimar a curva de sobrevivência, com resultados mais satisfatórios quando se lida com amostras pequenas.

O estimador Nelson-Aalen para a função  $l_x$ , obtido a partir do número de indivíduos em sucessivas idades  $i$ , anteriores a  $x$ ,

$$l_x = e^{-\sum_{i=i_{\text{inicial}}}^{x-1} \frac{D_i}{N_i}} \quad [7.12]$$

Sendo, como habitualmente,  $D_i$  o número de mortes com idade  $i$ . O somatório corre desde a idade inicial ( $i=i_{\text{inicial}}$ ) a partir da qual se segue o ciclo de vida, até à idade  $i=x-1$ .

<sup>2</sup> Este misterioso termo significa apenas que estes métodos não têm nenhum pressuposto *a priori* acerca da forma geral que a função  $l_x$  deve ter. O método Kaplan-Meier também é um método não-paramétrico. Num método paramétrico, adopta-se à partida uma determinada forma geométrica geral para  $l_x$ . Esta seria descrita por uma equação matemática com parâmetros que teriam de ser estimados a partir dos dados (em vez de se estimar idade a idade, como se faz nos não-paramétricos). Para uma descrição de métodos paramétricos, suas vantagens e inconvenientes, ver Skalski *et al* (2005) secção 5.3.4

A título exemplificativo, retomem-se os ursos. Acompanhemos o seu ciclo de vida a partir da idade inicial  $i=1,5$ , para a qual, por definição  $l_x=1$ . Vejamos o cálculo de  $l_x$  para as primeiras quatro idades,

$$l_{1,5} = 1$$

$$l_{2,5} = e^{-\left(\frac{2}{2}\right)} = 0.3679$$

$$l_{3,5} = e^{-\left(\frac{2}{2} + \frac{0}{7}\right)} = 0.3679$$

$$l_{4,5} = e^{-\left(\frac{2}{2} + \frac{0}{7} + \frac{2}{7}\right)} = 0.2765$$

$l_{1,5}$  é a probabilidade de um urso vivo no início do seguimento de uma coorte... estar vivo;  $l_{2,5}$  é a probabilidade de um urso vivo com 1.5 anos chegar vivo aos 2.5 anos. Agora, os dois ursos mortos entre 1.5 e 2.5 anos já não são problema. O método Nelson-Aalen fornece uma estimativa diferente de zero para  $l_{2,5}$ . A Tabela 2 é uma revisão das sobrevivências dos ursos, feita com o método Nelson-Aalen. Note-se que as sobrevivências estimadas,  $l_x$  são consideravelmente mais baixas no método Nelson-Aalen que no Kaplan-Meier, devido à incorporação da baixa sobrevivência dos jovens com 1.5 anos de idade.

Tabela 2. Taxas de sobrevivência até à idade ( $l_x$ ) dos ursos cinzentos, de acordo com o método Nelson-Aalen. A linha  $H_x$  é o expoente do estimador, i.e.,  $H_x = \sum_{i=1}^{x-1} D_i / N_i$ . Os ursos foram seguidos desde a idade  $i=1.5$ .

		Idades (anos)													
		1,5	2,5	3,5	4,5	5,5	6,5	7,5	8,5	9,5	10,5	11,5	12,5	13,5	14,5
<b><math>N_i</math></b>		2	7	7	7	7	6	6	9	8	9	5	2	3	2
<b><math>D_i</math></b>		2	0	2	1	1	0	1	0	0	1	0	0	0	0
<b><math>H_x</math></b>		0	1	1	1,286	1,429	1,571	1,571	1,738	1,738	1,738	1,849	1,849	1,849	1,849
<b><math>l_x</math></b>		1	0,3679	0,3679	0,2764	0,2395	0,2078	0,208	0,176	0,1759	0,1759	0,1574	0,157	0,1574	0,1574

		15,5	16,5	17,5	18,5	19,5	20,5	21,5	22,5	23,5	24,5	25,5	26,5	27,5	28,5
<b><math>N_i</math></b>		3	1	ND	1	1	2	2	2	ND	ND	2	2	1	1
<b><math>D_i</math></b>		1	0	ND	0	0	0	0	1	ND	ND	0	1	0	0
<b><math>H_x</math></b>		1,849	2,183	2,183	2,183	2,183	2,183	2,183	2,183	2,683	2,683	2,683	2,683	3,183	3,183
<b><math>l_x</math></b>		0,1574	0,1127	0,1127	0,1127	0,1127	0,1127	0,113	0,113	0,0684	0,0684	0,0684	0,068	0,0415	0,0415

A Figura 7.3 compara as curvas de sobrevivência pelos dois métodos.

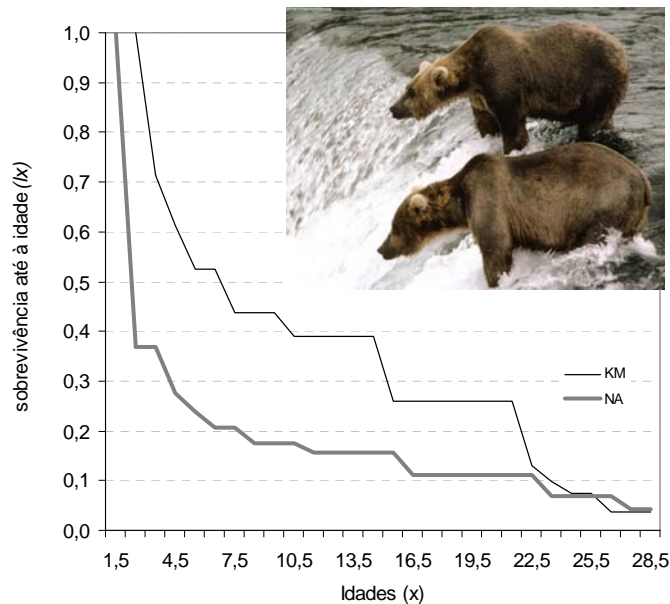


Figura 7.3. Curvas de sobrevivência ( $l_x$ ) dos ursos cinzentos pelo método Kaplan-Meier (KM) e Nelson-Aalen (NA).

## 7.6 Sobrevivência média

Em qualquer dos métodos apresentados para estimar  $S_x$ , as equações de referência para estimar sobrevivência são [4.1] e [4.2]. Os métodos (2) e (3) produzem com frequência mais do que uma estimativa da *mesma* taxa de sobrevivência, por exemplo,  $S_0$  pode ser estimado na passagem do ano  $t$  para  $t+1$  e, de novo, de  $t+1$  para  $t+2$ . O método (1) também pode produzir várias estimativas do mesmo  $S_x$ , uma por cada coorte seguida. Com várias estimativas do mesmo  $S_x$ , habitualmente o investigador pretende usá-las para estimar  $S_x$  médio. A forma correcta de obter esta média *não* é fazê-lo directamente a partir das várias estimativas de que se dispõe.

Em qualquer dos métodos, os valores de  $N_x$  usados para calcular a sobrevivência são, em geral, *estimativas* do número de indivíduos com  $x$  anos. São obtidos a partir de *sondagens* da população e não de recenseamentos. A diferença entre uma sondagem e um recenseamento (ou census) é que, na **sondagem**, é tomada uma *amostra* da população e, a partir desta, estimam-se os  $N_x$ . Num **recenseamento**, pelo contrário, todos os indivíduos são contados exhaustivamente e o valor de  $N_x$  é o valor real na população. Os valores de  $S_x$  e  $l_x$  calculados a partir de  $N_x$ 's estimados por sondagem são, eles próprios, também estimativas e, por isso, requerem que as suas propriedades estatísticas sejam tidas em atenção. John Gulland, um dos fundadores da moderna teoria de avaliação de recursos marinhos explorados pela pesca, estudou as propriedades estatísticas das equações [4.1] e [4.2] como estimadores de  $S_x$  e  $l_x$ . Gulland (1955) demonstrou, entre outras coisas, que [4.1] e [4.2] são estimadores não-centrados que tendem a sobreestimar o verdadeiro valor das taxas de sobrevivência na população. Por outras palavras, se o verdadeiro valor de uma taxa for  $S_x^v$ , a média de várias estimativas  $S_x^1, S_x^2, S_x^3, \dots$  tende a sobreestimar  $S_x^v$ .



Quando os  $N_x$  provêm de uma sondagem, existe um procedimento alternativo para estimar o valor médio de  $S_x$  e  $l_x$ , o qual é preferível à utilização directa de [4.1] e [4.2]. O procedimento correcto consiste em estimar primeiro as taxa instantâneas de mortalidade  $\mu_x$ , e depois usar as equações [4.11] e [4.14] para chegar a  $l_x$  e  $S_x$  a partir da média dos valores de  $\mu_x$ .

Sejam  $N_x$  e  $N_{x+\Delta x}$ , duas estimativas da abundância da população nas idades  $x$  e  $x+\Delta x$ , estimadas segundo um dos métodos acima descritos. Primeiro, estima-se a taxa  $\mu_x$  que se aplica a este periodo. Usando [4.14] e assumindo, para simplificar, que  $\Delta x=1$  ano,

$$\frac{N_{x+1}}{N_x} = e^{-\mu_x} \quad \text{logo} \quad \mu_x = -Ln \frac{N_{x+1}}{N_x} \quad [7.13]$$

Ao contrário de [4.1], a equação [7.13] é um estimador centrado de  $\mu_x$ . Se o valor inicial de  $N_x$  fôr  $N_0$ , e se se assumir que  $\mu_x$  se mantém constante desde 0 até  $x+\Delta x$ , a equação [4.11] permite calcular  $l_x$  a partir de  $\mu_x$ .

Obviamente, se a estimativa de  $S_x$  fôr feita a partir de um único par  $(N_x, N_{x+1})$ , a utilização de [7.13] seguido de [4.14], não apresenta qualquer vantagem relativamente a usar imediatamente [4.1], pois o resultado é rigorosamente o mesmo. É frequente, contudo, o biólogo encontrar-se na situação descrita em que dispõe de vários pares  $(N_x, N_{x+1})$  para o mesmo  $x$  e, portanto, de várias estimativas de  $\mu_x$ . O seguinte exemplo numérico ilustra este ponto.

Suponha-se que as seguintes *estimativas* de  $N_x$  ( $x= 0, 1, 2$ ) foram obtidas em quatro anos consecutivos (1995-98):

Idade (x)	1995	1996	1997	1998	
0	1000	1200	800	900	
1	800	600	900	400	
2	300	200	400	100	
$S_1$	0.250	0.667	0.111		$S_1$ médio: 0.343
$\mu_1$	1.386	0.405	2.197		$\mu_1$ médio 1.330

Os valores de  $S_1$  foram calculados usando a equação [4.1], para as três transições de anos (1995-96, 1996-97 e 1997-98) e estão assinalados na tabela acima, bem como a sua média, 0.343. Os valores de  $\mu_1$ , obtidos usando [7.13], e a respectiva média, também foram calculados. Suponhamos agora que se pretende estimar quantos indivíduos estarão vivos em 1999 com 2 anos de idade. Para isso precisamos de uma estimativa de  $S_1$ , a fim de efectuar a operação  $N_2=400 \times S_1$ . O mais natural seria adoptar como estimativa a média dos três valores de que já dispomos, isto é 0.343. Os resultados de Gulland (1955), porém, indicam que  $S_1=0.343$  sobreestima o verdadeiro valor de  $S_1$ . A alternativa possível, consiste em usar o valor médio  $\mu_1=1.330$  e, utilizando [4.14], obter a estimativa  $S_1=e^{-1.330} = 0.265$ . Esta estimativa de  $S_1$  é centrada e, portanto, preferível à anterior.

### 7.7 Sobrevivência total

Quanto à sobrevivência total da população,  $S_t$ , o mais frequente é ser estimada por um procedimento que é um caso particular do método (3) descrito na secção 7.2. Estima-se a abundância total em  $t$ , isto é  $N_t$ , e um intervalo de tempo  $\Delta t$  mais tarde, isto é  $N_{t+\Delta t}$ . O número de nascimentos  $B_t$ , ocorrido durante  $\Delta t$ , deve ser subtraído de  $N_{t+\Delta t}$ , pois os recém-nascidos não são sobreviventes de  $N_t$ . A sobrevivência  $S_t$  pode ser calculada aplicando directamente as equações [2.4] ou [2.5] no caso de reprodutores sazonais, ou [2.20] no caso de contínuos.

No caso de  $N_t$  e  $N_{t+\Delta t}$  serem estimativas da abundância e não o resultado de recenseamentos, estas equações também não são estimadores centrados e, por isso, caso se pretenda estimar a sobrevivência total média a partir de vários pares  $(N_t, N_{t+\Delta t})$ , aconselha-se de novo a metodologia proposta por Gulland (1955) na secção anterior. A ideia consiste em usar o logaritmo da equação [2.20] para estimar primeiro a taxa instantânea  $d$ ,

$$d\Delta t = -Ln \frac{N_{t+\Delta t} - B_t}{N_t} \quad [7.14]$$

e depois usar [2.21] para calcular a média de  $S_t$  a partir da média dos vários valores de  $d$ .

É possível também estimar taxas de sobrevivência médias para certos grupos etários, como um caso particular do método (3). Suponha-se que num dado instante se conhece  $N_{0+1+2}$ , a abundância dos indivíduos com 0, 1 e 2 anos de idade. Um ano mais tarde estima-se  $N_{1+2+3}$ , a abundância dos indivíduos com 1, 2 e 3 anos. Dividindo  $N_{1+2+3}$  por  $N_{0+1+2}$  estima-se a taxa de sobrevivência média de três coortes da população, neste caso a taxa de sobrevivência do grupo etário dos 0 aos 3 anos. O principal cuidado a ter consiste em não deixar que a segunda estimativa,  $N_{1+2+3}$ , seja contaminada por indivíduos com 0 anos. Estes últimos não eram ainda nascidos quando se efectuou a primeira estimativa e, portanto, não são sobreviventes da população então estimada.

### 7.8 Quanto tempo leva a estimar $N_x$ ?

Todas as equações que usámos para estimar sobrevivência, pressupõem que é possível efectuar medições *instantâneas* da abundância da população. Os valores de  $N_x$  e  $N_{x+\Delta x}$  nas equações, referem-se a instantes de tempo, um no início e outro no fim do intervalo  $\Delta x$ . Na prática, é impossível efectuar medições instantâneas da abundância. É costume, contudo, pressupor que o tempo que demora a obter as duas estimativas de  $N$  é negligível, relativamente ao intervalo de tempo que decorre entre elas. Suponha-se, por exemplo, que a abundância de cada idade é estimada todos os anos a partir de dados recolhidos num processo de amostragem que dura 15 dias, diga-se entre 1 e 15 de Janeiro. Em situações deste tipo, o período de amostragem parece suficientemente pequeno, relativamente ao intervalo de um ano, e parece razoável assumir que a estimação de  $N$  foi instantânea.

Por vezes, contudo, a sondagem da população é tão prolongada que a população está quase permanentemente a ser amostrada. Por exemplo, os biólogos das pescas estimam a abundância das populações de peixes com base nas capturas da pesca comercial efectuadas *ao longo do ano*. A pesca funciona portanto como um instrumento de sondagem que actua quase ininterruptamente. Por exemplo,  $N_x$  em 1999, representa o número de indivíduos com  $x$  anos ao longo de todo o ano de 1999, e não num instante apenas. Nestes casos, é melhor interpretar  $N_x$  como uma estimativa aproximada da

abundância média da idade  $x$  durante 1999, isto é  $N_{xm}$ . Apesar disso, mesmo nestes casos os biólogos costumam utilizar as equações e métodos das secções 4.1 e 4.2 (módulo 4) para estimar sobrevivência, como se  $N_{xm}$  fosse  $N_x$ . Qual é o pressuposto que se está a adoptar ao fazer isto ? Ou seja, em que condições é que  $S_x = N_{x+1}/N_x$  pode ser substituído por  $S_x = N_{(x+1)m}/N_{xm}$  ? ( $N_{(x+1)m}$  = número médio de indivíduos com idade  $x+1$ ). Para responder, vou examinar este último quociente um pouco melhor, usando [4.15]:

$$\frac{N_{(x+1)m}}{N_{xm}} = \frac{\frac{N_{x+1}(1 - e^{-\mu_{x+1}})}{\mu_{x+1}}}{\frac{N_x(1 - e^{-\mu_x})}{\mu_x}} = S_x \frac{\mu_x}{\mu_{x+1}} \frac{1 - e^{-\mu_{x+1}}}{1 - e^{-\mu_x}} \quad [7.15]$$

O lado direito de [7.15] mostra que  $N_{(x+1)m}/N_{xm}$  só é igual a  $S_x$  se as taxas de mortalidade que actuam sobre as idades  $x$  e  $x+1$  forem iguais, isto é, se  $\mu_x = \mu_{x+1}$ . Se os valores destas taxas forem muito pequenos,  $N_{(x+1)m}/N_{xm}$  também constitui uma boa aproximação a  $S_x$ .

## Literatura Citada

- Aalen, O. 1978. Nonparametric inference for a family of counting processes. *Annals of Statistics* **6**:701-726.
- Agresti, A, and BA Coull. 1998. Approximate is better than "exact" for interval estimation of binomial proportions. *Amer Statistician* **52**(2): 119-126
- Brown, LD, TT Cai, and A DasGupta. 2001. Interval estimation for a binomial proportion. *Statistical Science* **16**(2): 101-133
- Carey, J.R. 1993. *Applied Demography for Biologists*. Oxford Univ. Press, Oxford.
- Caswell, H. 1989. *Matrix Population Models*. Sinauer, Sunderland, Mass.
- Cochran, W. 1977 (3<sup>rd</sup> ed). *Sampling Techniques*. John Wiley.
- Dempster, JR. 1975. *Animal Population Ecology*, Academic Press, New York.
- Ebert, T.A. 1999. *Plant and Animal Populations. Methods in Demography*. Academic Press, San Diego, Calif.
- Gulland, J.A. 1955. Estimation of growth and mortality in commercial fish populations. *U.K. Min. Agric. Fish., Fishery Invest. (Ser. 2)* **18**(9): 46 pp.
- Gulland, J.A. 1969. *Manual of Methods for Fish Stock Assessment. Part 1. Fish Population Analysis*. Food and Agriculture Organization of the UN, Man. Fish. Sci. **4**:1-154.
- Hoening, J.M. 1984. Empirical use of longevity data to estimate mortality rates. *Fishery Bulletin* **81**(4).

- Hoenig, J.M., W.D. Lawing, and N.A. Hoenig. 1983. Using mean age, mean length, and median length data to estimate the total mortality rate. *Int. Counc. Explor. Sea. C.M.* **23**:1-11.
- INE. 1997. *Estimativas de População Residente*, Instituto Nacional de Estatística, Lisboa.
- INE. 1998. *Estatísticas Demográficas*, Instituto Nacional de Estatística, Lisboa.
- Keyfitz, N and H Caswell. 2005, 3<sup>rd</sup> Ed. *Applied Mathematical Demography*. Springer Science, NY.
- King M. 1995. *Fisheries Biology, Assessment and Management*. Fishing News Books, Blackwell Sci., London.
- Klein, JP and ML Moeschberger. 1997. *Survival Analysis. Techniques for Censored and Truncated Data*. Springer-Verlag, NY.
- May RM. 1976. Estimating r: A pedagogical note. *American Naturalist* **110**:496-499.
- McCallum H. 2000. *Population Parameters: Estimation for Ecological Models*. Blackwell Science Ltd
- Michod RE and Anderson WW. 1980. On calculating demographic parameters from age frequency data. *Ecology* **61**:265-269.
- Nelson, W. 1972. Theory and applications of hazard plotting for censored failure data. *Technometrics* **14**:946-965.
- Pauly, D. 1984. *Fish Population Dynamics in Tropical Waters: a Manual for Use with Programmable Calculators*. ICLARM, Manila.
- Peters, RH. 1983. *The Ecological Implications of Body Size*. Cambridge Univ. Press, Cambridge.
- Pianka ER. 1978. 2<sup>nd</sup> Ed. *Evolutionary Ecology*. Harper & Row, NY.
- Pollock, KH, SR Winterstein, CM Bunck, and PD Curtis. 1989. Survival analysis in telemetry studies: The staggered entry design. *Journal of Wildlife Management* **53**:7-15.
- Pollard JH. 1973. *Mathematical Models for the Growth of Human Populations*. Cambridge Univ. Press, Cambridge.
- Ricker, W.E. 1975. *Computation and Interpretation of Biological Statistics of Fish Populations*. Fisheries Research Board of Canada, Bull. 191.
- Roughgarden J. 1979. *Theory of Population Genetics and Evolutionary Ecology: An Introduction*. MacMillan Pub., NY.
- Skalski JR, KE Ryding, and JJ Millspaugh. 2005. *Wildlife Demography. Analysis of Sex, Age, and Count Data*. Elsevier, Academic Press, Amsterdam.
- Southwood TRE, 1978. *Ecological Methods*. Chapman and Hall, London.
- Yodzis P. 1989. *Introduction to Theoretical Ecology*. Harper & Row Publ., NY.
- Zar, JH. *Biostatistical Analysis*, 2<sup>nd</sup> Ed. Prentice Hall, New Jersey.