

Módulo 4

Demografia: a coorte

A unidade básica da análise demográfica é o *indivíduo*. Os indivíduos nascem, crescem, reproduzem-se e morrem. As taxas com que estes processos individuais decorrem a nível da população, vulgarmente designadas por *taxas vitais*, determinam se a população aumenta, diminui, oscila ou mantém-se estável. O meio ambiente afecta a população através da sua acção sobre as taxas vitais e, para compreender a resposta da população ao ambiente, é necessário integrar a informação sobre as taxas vitais e ligá-la aos processos populacionais. Ao fazer esta integração, é necessário ter em conta as diferenças que existem entre os próprios indivíduos. Provavelmente, a forma mais fácil de o fazer consiste em organizar os indivíduos (i.e. “estruturar” a população) em grupos que tenham taxas vitais relativamente parecidas. Normalmente, estes grupos reúnem indivíduos em estádios de desenvolvimento fisiológico idêntico. O exemplo mais comum consiste em agrupar os indivíduos por idades dentro da população e, nessas circunstâncias, fala-se da *estrutura etária da população*.

A estrutura etária da população é objecto de estudo da **Demografia**. Historicamente a demografia desenvolveu-se muito centrada na população humana, nomeadamente no crescimento exponencial da população e no cálculo das probabilidades de sobrevivência, motivados por razões económicas ou sociais. Aparentemente, no século III os romanos já tentavam medir a variação da mortalidade com a idade e este esforço foi muito mais tarde revitalizado com o aparecimento do chamado Cálculo Atuarial, conduzido pelas companhias de seguros para estimar os prémios dos seguros de vida. Raymond Pearl foi um dos primeiros ecologistas que, reconhecendo a possível utilidade dos estudos atuariais para os ecologistas, introduziu, na década de 1920, métodos demográficos no estudo de populações não-humanas.

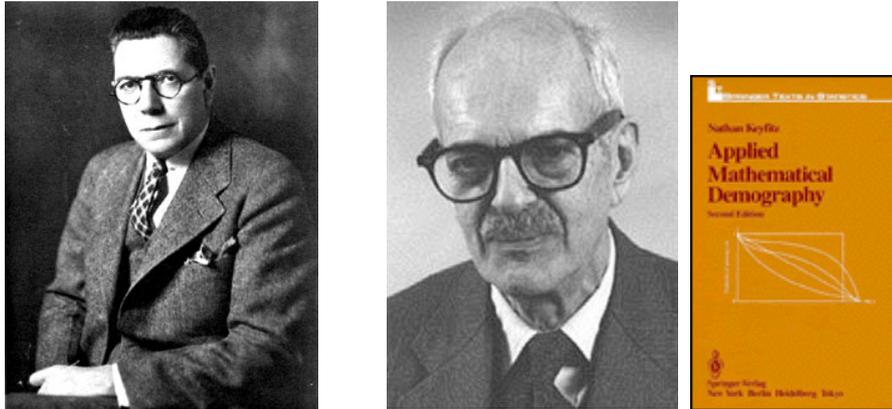


Figura 4.1 Esquerda - Ray Pearl (1879-1940) fez parte dum grupo de biólogos que, na viragem do séc 19 para 20, procurou libertar-se da biologia essencialmente descritiva, baseada na descrição morfológica de animais e plantas, que marcou o séc 19. Pearl foi um dos fundadores da ecologia populacional e procurou leis gerais, transversais aos vários taxa. Foi também um dos percussores da generalização de ferramentas matemáticas em biologia. Pearl é conhecido pela sua defesa da curva logística como uma 'lei' geral de crescimento que se aplica também a populações humanas, facto que lhe granjeou muitos críticos.

Direita – Natham Keyfitz (1913-2009) provavelmente o maior demógrafo do século 20. Licenciou-se em matemática, começou por trabalhar no Bureau of Statistics do Canadá e é creditado como o principal pioneiro no desenvolvimento da moderna demografia matemática. É autor de centenas de livros e artigos.

Roughgarden refere-se à demografia como sendo “The study of population growth with age structure” (Roughgarden 1979 p. 323), uma definição que parece ainda muito ligada ao peso que a população humana teve no desenvolvimento dos métodos demográficos. Na verdade, a moderna aplicação de métodos demográficos em ecologia é muito menos restritiva, pois admite que os indivíduos da população possam ser agrupados segundo critérios diferentes da idade (Caswell 2001, Tuljapurkar and Caswell 1997). Exemplos de outros critérios, são o tamanho corporal dos indivíduos ou o estágio do ciclo de vida em que se encontram. Modernamente, dispomos de uma teoria demográfica de populações estruturadas segundo critérios fisiológicos, entre os quais, a idade, é apenas um caso particular. Contudo, esta teoria demográfica mais geral de populações estruturadas é menos familiar e matematicamente mais complicada. A sua aplicação prática para projectar o futuro de populações levanta também dificuldades experimentais maiores. Assim sendo, a definição de Roughgarden é provavelmente um bom ponto de partida para a iniciação à demografia e, por isso, será essencialmente com as populações estruturadas por idade que nos iremos ocupar nesta fase inicial.

Embora a demografia seja uma área do conhecimento com especificidade própria, no mundo há muito poucos departamentos académicos, ou mesmo grupos de estudos graduados, devotados exclusivamente à demografia. Se pusermos de lado o pessoal encarregue de recolher e processar dados de recenseamentos humanos, a grande maioria dos investigadores e praticantes

dos métodos demográficos têm em geral os seus interesses mais centrados noutras áreas. É o caso da sociologia e economia, no caso restrito das populações humanas, e da ecologia, epidemiologia, pescas, caça, reservas naturais, florestas e entomologia, no caso mais geral. Apesar disso, a teoria demográfica, na sua expressão geral, é igualmente válida e aplicável em todas estas áreas de investigação. Na prática, porém, cada área tem criado a sua própria terminologia e desenvolvido as técnicas mais apropriadas às suas necessidades.

A unidade demográfica imediatamente acima do indivíduo é a coorte. Uma **coorte** (ou **classe anual**) é o conjunto de todos os indivíduos nascidos no mesmo intervalo de tempo (ano, mês...) *ao longo do tempo, até haver pelo menos um deles vivo*. Por exemplo, o conjunto de javalis nascidos na tapada de Mafra durante 2005, deu início à coorte de 2005 de javalis. Possuíam 0 anos de idade em 2005, fizeram 1 ano de idade em 2006, fizeram 2 anos em 2007, ..., farão 5 anos em 2010 etc.. Estes que têm 5 anos em 2010, evidentemente, são apenas uma fracção de todos os que nasceram em 2005, são os chamados *sobreviventes da coorte* de 2005. De facto, com o passar do tempo, o número de indivíduos de uma dada coorte *só pode diminuir*, mas nem por isso deixam de manter “viva” a coorte a que pertencem. A tabela a seguir ilustra a coorte de 1997 de uma população ao longo de 6 anos da sua vida,

TABELA 4.1. Evolução do número de indivíduos, N_x , da coorte de 1997, ao longo dos anos.

<i>Idades</i>	<i>Anos do calendário</i>					
	1997	1998	1999	2000	2001	2002
0	1000					
1		800				
2			600			
3				450		
4					300	
5						150

Como ficou implícito acima, a coorte é designada pelo ano do calendário em que teve início. Fala-se assim da coorte de 1997, da coorte de 1999 etc. Em populações aquáticas, é comum designar o número de indivíduos com que a coorte se inicia por **recrutamento** à coorte. Tem-se assim anos (ou coortes) de forte recrutamento, de recrutamento médio etc.. No quadro acima, a coorte de 1997 teve um recrutamento de 1000 indivíduos. Note-se, finalmente, que o conceito de coorte tem em si qualquer coisa de abstracto, uma vez que uma coorte não existe totalmente num dado *instante* de tempo mas sim ao longo de um *intervalo* de tempo. A população, pelo contrário, não tem nada de abstracto. Existe num determinado instante de tempo e é composta por *representantes de várias coortes*, tantas quanto o número de idades presentes na população.



Curiosidade histórica – A coorte era a unidade tática básica da legião romana. Uma legião (cerca de 5400 homens) era formada por 10 coortes, designadas simplesmente, 1ª coorte, 2ª coorte etc, sendo a 1ª a dos legionários mais antigos e prestigiados. Uma coorte era formada por seis ‘centuria’ de 80 homens, cada uma comandada por um centurião. O centurião mais sênior também comandava a coorte inteira. Na imagem, uma coorte em manobras.

4.1 Variação do número de indivíduos da coorte em tempo discreto

Vou usar a letra x para representar as idades. Suponhamos que recenseamos a abundância da coorte em duas idades sucessivas x e $x+1$ separadas pela unidade de tempo (e de idade) Δx . N_x é o número de indivíduos com x anos de idade e N_{x+1} é o número de indivíduos após Δx , sendo evidentemente $N_x \geq N_{x+1}$. Define-se **taxa de sobrevivência entre as idades x e $x+1$** , e representa-se por S_x , a seguinte quantidade:

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

Quando se utiliza S_x , é habitual utilizar números inteiros para a idade, $x=0, 1, 2, \dots$, e pressupor que x se reporta ao *início* da idade. Por exemplo, S_2 é a sobrevivência dos indivíduos com 2 anos de idade. Se a unidade de tempo adoptada for 1 ano, S_2 aplica-se entre o instante em que os indivíduos fazem 2 anos e o instante em que fazem 3 anos. S_x pode ser interpretada como a probabilidade de que um indivíduo que atinge a idade x sobreviva durante mais uma unidade de tempo. Trata-se portanto de uma probabilidade de sobrevivência *condicional* a que o indivíduo esteja vivo no início da idade x .

Exercício Assumindo que os números da coorte na Tabela 4.1 representam o número de

indivíduos no *inicio* de cada uma das idades x , verificar que $S_0=0.8$, $S_1=0.75$, $S_4=0.5$.

Define-se l_x como a **taxa de sobrevivência até à idade x** , isto é, a probabilidade de que um recém-nascido chegue vivo até ao *inicio* da idade x . Matematicamente,

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

Note-se que $l_0=1$, i.e. a probabilidade de um recém-nascido iniciar a idade de 0 anos é 1.

Exercício Com o mesmo pressuposto do exercício anterior, verificar que $l_4=0.3$ e $l_5=0.15$

A inspeção de [4.1] e [4.2] sugere logo a existência de relações entre S_x e l_x . A primeira delas é,

$$l_x = S_0 S_1 S_2 \dots S_{x-1} \quad [4.3]$$

A demonstração de [4.3] no caso l_3 , por exemplo, é,

$$l_3 = S_0 S_1 S_2 = \frac{N_1}{N_0} \frac{N_2}{N_1} \frac{N_3}{N_2} = \frac{N_3}{N_0}$$

A segunda relação é,

$$S_x = \frac{l_{x+1}}{l_x} \quad [4.4]$$

Exercício Demonstrar [4.4].

Sempre que se tem uma taxa de sobrevivência, é possível convertê-la em taxa de mortalidade e vice-versa. De facto, define-se a **taxa de mortalidade entre as idades x e $x+1$** , q_x , simplesmente como:

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

sendo $D_x=N_x-N_{x+1}$ o **número de mortes** da coorte com x anos de idade. À semelhança de S_x , também q_x pode ser entendido como a probabilidade de um indivíduo morrer no intervalo de idades que vai de x a $x+1$, condicional a que o indivíduo tenha chegado vivo à idade x .

4.2 Variação do número de indivíduos em tempo contínuo

Até aqui, defini taxas vitais para uma coorte em termos da variação do seu número de indivíduos nos intervalos Δx , entre as idades discretas $x= 0, 1, 2, \dots$. A variação do número de indivíduos de uma coorte, contudo, deve-se apenas à mortalidade (assumindo ausência de migração) e, na maioria das populações, a mortalidade ocorre *contínuamente* ao longo do tempo. Assim, pelos mesmos argumentos usados na Secção 2.2 (módulo 2), quando passei dos reprodutores sazonais para os contínuos, é mais natural falar na *variação instantânea* do número de indivíduos da coorte do que de variação “aos saltos” entre idades discretas.

O conceito de variação média do número de indivíduos, definido quando discuti o crescimento da população (equação [2.1]), também se aplica aos indivíduos de uma coorte no intervalo Δx . Seguindo um raciocínio idêntico ao que levou à equação [2.14], a variação instantânea de N_x obtém-se quando se considera a variação média de N_x dentro do intervalo Δx e este intervalo tende para zero (recordar [2.11]). No caso da coorte, a única causa de variação no instante x , é o número de mortes que ocorre nesse instante, D_x , i.e.,

$$\frac{dN}{dx} = -D_x \quad [4.6]$$

Esta variação é negativa ($N_{x+\Delta x} - N_x < 0$), daí D_x estar afectado de um sinal negativo.

Se $\mu_x = D_x/N_x$ representar a **taxa instantânea de mortalidade da idade x** , então,

$$\frac{dN}{dx} = -\mu_x N_x \quad [4.7]$$

Note-se que μ_x pode ser, ou não, igual à taxa instantânea d , definida para todas as idades da população em simultâneo (eq [2.13]). Se o valor de μ_x for igual para todas as idades, então $d = \mu_x$. A solução de [4.7] obtém-se pelo mesmo método que foi usado na Secção 2.2 para obter [2.17], e é:

$$N_{x+\Delta x} = N_x e^{\int_x^{x+\Delta x} -\mu_x dx} \quad [4.8]$$

Se assumirmos que μ_x se mantém constante entre as idades x e $x+\Delta x$, um pressuposto que parece razoável para um Δx não muito grande, então [4.8] simplifica-se:

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

Se conhecermos o número de indivíduos na idade x e a taxa μ , a equação [4.9] permite conhecer quantos indivíduos ainda estarão presentes na coorte na idade $x+\Delta x$, sob o pressuposto de μ constante. A solução [4.9] mostra que o decréscimo do número de indivíduos dá-se de forma exponencial negativa. No caso particular em que x é a idade 0 e, portanto, N_0 são os recém-nascidos que originam a coorte, então, pressupondo que a coorte está sujeita a uma taxa μ constante, o número de indivíduos ainda presentes com x anos de idade é,

$$N_x = N_0 e^{-\mu x} \quad [4.10]$$

A *proporção* de indivíduos que ainda estão vivos na idade x , relativamente aos N_0 que nasceram, já foi definida pelas equações [4.2] a [4.3]. Em termos mais rigorosos (ainda assumindo μ constante) l_x é simplesmente,

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

O número total de indivíduos que compõem a coorte, N , é a soma acumulada dos indivíduos de todas as idades,

$$N = N_0 + N_1 + N_2 + \dots = \sum_{x=0}^{\infty} N_x \quad [4.12]$$

A equação [4.12] é útil na prática, contudo, como a idade é uma variável contínua, a forma correcta de calcular N consiste em integrar [4.9] a partir da idade $x=0$. No caso particular em que μ_x é constante e diferente de zero, a equação a integrar é a [4.10]:

$$N = \int_0^{\infty} N_0 e^{-\mu x} dx = N_0 \left(\frac{e^{-\infty}}{-\mu} - \frac{e^0}{-\mu} \right) = \frac{N_0}{\mu} \quad [4.13]$$

O número total de indivíduos da coorte é igual ao número total de indivíduos que inicia a coorte a multiplicar pelo inverso da taxa instantânea de mortalidade.

As equações [4.1] e [4.9], sugerem que existe uma relação entre a taxa instantânea de mortalidade da idade x e a respectiva taxa de sobrevivência. Se Δx for o intervalo de tempo unitário que queremos usar (por exemplo $\Delta x=1$ ano), e se designarmos por μ_x a taxa instantânea de mortalidade constante que actua dentro de Δx , então de [4.1] e [4.9] obtem-se,

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

As equações [4.8] a [4.10] permitem calcular o número de indivíduos na coorte em qualquer idade. Na prática, porém, raramente se conhece a abundância de uma coorte em instantes exactos de tempo; é mais comum ter estimativas da abundância ao longo de intervalos finitos, como por exemplo um ano. É útil portanto conhecer o número *médio* de indivíduos entre duas idades, x e $x+\Delta x$, simbólicamente N_{xm} . Pretende-se, por exemplo, saber quantos indivíduos houve em média com 2 anos de idade, e não quantos houve nos instantes em que a coorte fez 2 ou 3 anos de idade. Seria possível calcular simplesmente a média aritmética, $N_{xm}=(N_x+N_{x+\Delta x})/2$, contudo, como vimos, o decréscimo de N_x para $N_{x+\Delta x}$ não é linear mas sim exponencial, pelo que a média aritmética sobre-estimaria o verdadeiro valor médio. O cálculo de N_{xm} , dentro de um intervalo Δx , entre a idade que se toma para início da contagem do tempo, $x=0$, e o fim do intervalo, é feito integrando [4.9] e dividindo pelo próprio intervalo Δx :

$$N_{xm} = \frac{1}{\Delta x} \int_0^{\Delta x} N_x e^{-\mu x'} dx' = \frac{1}{\Delta x} N_x \left(\frac{e^{-\mu \Delta x} - e^0}{-\mu} - \frac{e^0}{-\mu} \right)$$

Assumindo que Δx é a nossa unidade básica de tempo ($\Delta x=1$), o **número médio de indivíduos** entre as idades x e $x+1$ é,

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

sendo μ_x a taxa instantânea de mortalidade, assumida constante, que actua dentro do intervalo de idade Δx .

4.3 Curvas de sobrevivência

A **curva de sobrevivência** ou função de sobrevivência é simplesmente a proporção de indivíduos que sobrevivem em função da idade. Os ecologistas costumam traçar a curva de sobrevivência em geral com o objectivo de discutir a forma como a morte actua ao longo do ciclo de vida de uma espécie. Para que a curva possa ser usada para fins comparativos entre espécies, deve ser traçada em unidades independentes do número absoluto de indivíduos numa coorte. Habitualmente a curva é o gráfico de l_x contra x e, portanto, indica a proporção de recém-nascidos

que ainda estão vivos em cada idade. Pode também ser interpretada como a variação da probabilidade dum recém-nascido chegar à idade x .

Desde Pearl (1927) e Deevey (1947) que os ecologistas distinguem três estereótipos de curvas de sobrevivência (I, II, III) que represento na Figura 4.2. O tipo I é tipificado pelos grandes mamíferos, incluindo a população humana moderna, alguns invertebrados marinhos, excluindo a fase de vida planctónica, e muitas plantas superiores. A sobrevivência é muito elevada na primeira parte do ciclo de vida e, provavelmente, é a curva menos comum na natureza. Mesmo para populações humanas, só há relativamente poucos séculos se aplica. A curva assume que uma percentagem muito alta da população tem uma longevidade que se aproxima muito da longevidade máxima e depois na “velhice” todos começam a morrer rapidamente. A curva II implica uma taxa de sobrevivência constante, independente da idade, e é típica da fase pós-juvenil de muitas aves, mamíferos roedores e algumas plantas. A curva III é típica de muitas espécies marinhas, muitos insectos e parasitas, com taxas de sobrevivência baixíssimas nas idades jovens. A vasta maioria das populações biológicas aproxima-se do tipo III. Na prática, contudo, as curvas de populações reais caem em posições intermédias entre as curvas protótipo e têm maiores irregularidades do que estas.

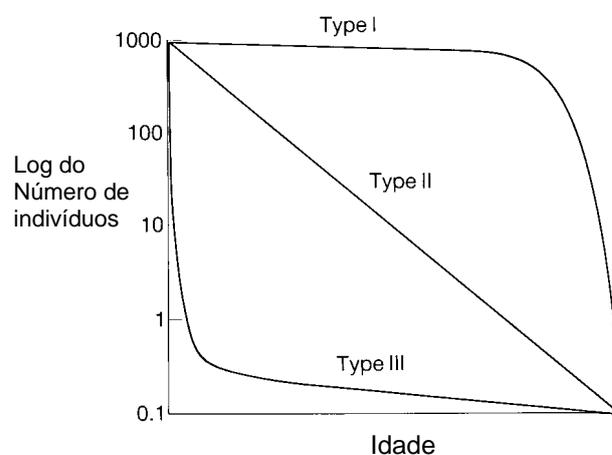


Figura 4.2. Curvas de sobrevivência estereotipadas, Tipos I, II e III de Deevey (1947). Em ordenadas, a escala logarítmica é adequada para representar *percentagem* de variação, em vez de variação absoluta. Por exemplo, considere-se duas situações em que os indivíduos da coorte foram reduzidos a 50%. Num caso passaram de 1000 para 500 indivíduos (variação absoluta=500), noutra de 500 para 250 indivíduos (variação=250 indivíduos). Em escala logarítmica (base 10) estes dois decréscimos são iguais: $\log 1000 - \log 500 = \log 500 - \log 250$, isto é, $3.00 - 2.70 = 2.70 - 2.40$, ou ainda, $0.3 = 0.3$. Os números de mortes são muito diferentes, mas as taxas de sobrevivência são iguais. Torna-se assim possível a comparação da variação do número de indivíduos em diferentes fases do ciclo de vida.

As curvas de sobrevivência são funções contínuas e, para as traçar, das duas uma. Ou um grupo de indivíduos é marcado individualmente e, em cada amostragem, o número de sobreviventes marcados pode ser contabilizado (um procedimento comum com plantas ou animais sésseis) ou, se não podem ser acompanhados individualmente, pode-se pelo menos acompanhar uma coorte e estimar o número de indivíduos presentes em cada momento de amostragem. A principal vantagem das curvas de sobrevivência é que, para as traçar, não é necessário obter estimativas da sobrevivência de todas as idades. Com um número de pontos no gráfico muito inferior ao número de idades já é possível traçar a curva. Esta vantagem, contudo, pressupõe que a curva de sobrevivência segue um modelo matemático pré-estabelecido. Por exemplo, se se assumir que a sobrevivência é do Tipo II em escala logarítmica (Fig. 4.2), o modelo é uma linha recta com inclinação negativa e bastariam dois pontos no gráfico para determinar a equação da recta. Claro que quanto mais pontos melhor, mas mesmo assim o número de pontos pode ficar muito aquém do número de idades. Evidentemente, o pressuposto de que existe um modelo geométrico subjacente à curva é muito forte e pode falhar.

Para ajustar uma curva de sobrevivência a pontos no gráfico, é necessário começar por seleccionar a função matemática que se julga apropriada para descrever a curva de sobrevivência e, em seguida, estimar os seus parâmetros. Há duas funções muito usadas em ecologia e que em geral são suficientes (McCallum 2000) para quase todas as situações: a função exponencial e a função de Weibull. A função exponencial parte do pressuposto que a taxa instantânea de mortalidade é constante com a idade e, portanto, a variação do número de indivíduos da coorte é dada pela nossa equação [4.7], sendo $\mu_x = \mu$ uma constante,

$$\frac{dN}{dx} = -\mu N_x \quad [4.7]$$

A curva de sobrevivência até à idade x é então a função [4.11], na qual a sobrevivência declina exponencialmente:

$$l_x = e^{-\mu x} \quad \text{i.e.} \quad \ln l_x = -\mu x \quad [4.16]$$

A equação [4.16] é uma recta (curva Tipo II) cuja inclinação é $-\mu$, a qual pode ser estimada a partir de um conjunto de pontos.

O caso da curva Tipo II é o mais simples, pois como se viu só há um parâmetro a estimar – o declive da recta (a ordenada na origem localiza-se obrigatoriamente no número inicial de indivíduos). Para curvas mais realistas, contudo, a função deve ser mais complexa e é habitual recorrer à função de Weibull. Esta função também tem a taxa instantânea μ e,

adicionalmente, tem um parâmetro sem dimensões que lhe confere plasticidade, θ . Matematicamente:

$$l_x = e^{(-\mu x)^\theta} \quad \text{i.e.} \quad \text{Ln } l_x = (-\mu x)^\theta \quad [4.17]$$

O parâmetro θ determina se a taxa instantânea de mortalidade aumenta ou diminui ao longo de um intervalo de idades. Se $\theta > 1$, gera-se uma mortalidade que aumenta com a idade (curva Tipo III) e se $\theta < 1$ a mortalidade diminui com a idade (curva Tipo I). Se $\theta = 1$, a função de Weibull reduz-se à função exponencial. Para ajustar um conjunto de pontos à função de Weibull, pode-se usar um pacote estatístico comercial, mas também é possível recorrer a uma simples folha de cálculo (e.g. Excel) e usar princípios básicos do método da máxima verosimilhança para estimar μ e θ (McCallum 2000, Cap 4, por exemplo, explica a receita a usar).

Literatura Citada

Deevey, E.S. 1947. Life tables for natural populations of animals. *Quarterly Review of Biology* **22**:283-314.

McCallum H. 2000. *Population Parameters: Estimation for Ecological Models*. Blackwell Science Ltd.

Roughgarden, J. 1979. *Theory of Population Genetics and Evolutionary Ecology: An Introduction*. MacMillan Publ, NY.