CAPÍTULO 7

CARACTERÍSTICAS, EVENTOS, PROPORÇÕES, TAXAS E PROBABILIDADES

7.1 RAZÕES, PROPORÇÕES, TAXAS E ÍNDICES

Entre outras coisas, este capítulo introduzirá o conceito de *taxas*. Antes de entrar propriamente na aplicação deste conceito, convém refletir por um momento sobre o uso da palavra "taxa" e sua diferença de outros termos parecidos como "proporção", "razão", "índice" e "probabilidade". O que estes termos têm em comum é que geralmente são o resultado de uma divisão entre dois números. Entretanto, os demógrafos tendem a ser bastante cuidadosos com o uso apropriado de cada termo, dependendo daquilo que está no numerador e no denominador. A Razão de Sexos, que foi introduzida no capítulo anterior, é uma razão e não uma proporção porque o numerador não faz parte do denominador. Também seria possível (embora pouco usual) definir a proporção de homens ou de mulheres, dividindo o número de homens ou mulheres pela população total. Já que tanto os homens como as mulheres fazem parte da população total, estas efetivamente são proporções. Um outro exemplo é o grau de urbanização de uma população, ou seja, o número de habitantes urbanos dividido pela população total (eventualmente vezes 100). Embora não seja o termo usual, esta quantidade pode ser legitimamente chamada a "proporção de urbanização". Também não seria incorreto chamá-la de "razão de urbanização" porque cada proporção é também uma razão. Mas é definitivamente incorreto chamá-la de "taxa de urbanização", como se faz frequentemente.

Como se verá mais adiante, as taxas e as probabilidades são quantidades dinâmicas que envolvem tanto variáveis de fluxo como de estoque. Este não é o caso nem do grau de urbanização nem da percentagem de pessoas não alfabetizadas numa população, muitas vezes erroneamente chamada de "taxa de analfabetismo", ou da proporção de desemprego, que geralmente é chamada "taxa de desemprego", embora formalmente não seja uma taxa. Na demografia, para que uma quantidade seja chamada uma "taxa", ela precisa ter um numerador que quantifica um número de eventos e um denominador que descreve o número de pessoas que em alguma medida podem ser expostas a este evento. Ainda se distingue entre taxas mais puras, nas quais todas ou a grande maioria das pessoas no denominador têm uma possibilidade real de experimentar o evento, e taxas mais brutas, cujo denominador inclui muitas pessoas que em realidade não estão expostas ao evento. Por exemplo, a Taxa Bruta de Natalidade (ver seção 10.1 do Capítulo 10) inclui homens no seu denominador enquanto as taxas de fecundidade mais puras se baseiam inteiramente nas mulheres. Infelizmente, o uso da terminologia nem sempre é consistente, nem mesmo dentro da demografia. Por exemplo, o número médio de filhos nascidos vivos que as mulheres têm ao longo das suas vidas na literatura norte-americana e brasileira geralmente é chamada a Taxa de Fecundidade Total (ver Capítulo 10), embora segundo o critério anterior não se trate de uma taxa. Na literatura em francês e em Portugal a Taxa de Fecundidade Total é geralmente chamada o Índice Sintético de Fecundidade ou Descendência Final, que são termos mais corretos.

Finalmente ainda existe o termo "índice" que é o mais geral de todos. Qualquer número que quantifica uma relação observada no mundo natural ou social pode ser considerado um "índice". Neste sentido não seria errado falar do "Índice de Sexos", embora este não seja o termo habitual.

7.2 O CRESCIMENTO DA POPULAÇÃO

As populações e as suas componentes mudam ao longo do tempo. Esta mudança pode ser tanto positiva como negativa (para mais e para menos) e em ambos os casos se usa o termo crescimento da população. Como já se mencionou no Capítulo 2, o mundo está saindo de um período histórico no qual o crescimento era quase universalmente positivo para um período de maior variedade de situações em que um número considerável de países está experimentando crescimento negativo das suas populações, situação esta que já existe em Portugal e que nos próximos 25-30 anos também acontecerá no Brasil. A próxima seção indagará sobre as diferentes componentes do crescimento. Mas antes disso é preciso dizer algumas palavras sobre como se caracteriza o ritmo de crescimento de uma população mais em geral.

O ponto de partida para qualquer medição do crescimento é a comparação do tamanho de uma população ou subpopulação em dois momentos do tempo, t e $t+\Delta t$. Para certos propósitos é suficiente comparar os tamanhos absolutos nestes dois momentos e calcular uma diferença. Mas para outros propósitos é preciso relacionar este aumento com o tamanho inicial da população. Afinal, um crescimento de 1.000 indivíduos numa aldeia de 500 habitantes tem implicações totalmente diferentes do mesmo crescimento numa metrópole de 10 milhões. Portanto, muitas vezes se calcula o crescimento em termos relativos: $P(t+\Delta t) / P(t)$. Mas isso ainda deixa a dúvida como relacionar esta razão com o tempo Δt. Para padronizar a taxa, de modo que não dependa diretamente de Δt, usam se dois conceitos: o de *taxa anual* e *taxa instantânea* ou *contínua*, com as seguintes definições:

$$\text{Anual: } \quad r = (P(t+\Delta t) / P(t))^{1/\Delta t} - 1 \tag{7.1.a}$$

$$\text{Instantânea: } \quad r = \ln(P(t+\Delta t) / P(t)) / \Delta t \tag{7.1.b}$$

O símbolo *ln* se refere ao logaritmo natural ou de base $e = 2{,}71828$...... A contraparte das fórmulas (7.1.a-b) no cálculo de $P(t+\Delta t)$ a partir de $P(t)$, usando r, é a seguinte:

$$\text{Anual:} \qquad P(t+\Delta t) = P(t) \cdot (1+r)^{\Delta t} \tag{7.2.a}$$

$$\text{Instantânea:} \qquad P(t+\Delta t) = P(t) \cdot e^{r\Delta t} \tag{7.2.b}$$

Os dois conceitos de r se relacionam da seguinte forma:

$$r_{inst.} = ln(1 + r_{anual}) \tag{7.3}$$

A razão porque existem estes dois conceitos distintos é que existem as duas fórmulas distintas (7.2.a) e (7.2.b) para a projeção do crescimento que são identificadas às vezes como a fórmula *geométrica* e a fórmula *exponencial*, respectivamente. Por esta razão, a taxa anual também é conhecida como taxa geométrica e a instantânea como taxa exponencial de crescimento. Por enquanto será usado só o conceito de crescimento anual (7.1.a), em combinação com a fórmula (7.2.a). De qualquer modo, a diferença entre as duas variantes de r é pequena. Por exemplo, segundo a Tabela 2.1 do Capítulo 2, se espera que a população da América Latina aumentará de 634 milhões em 2015 para 784 milhões em 2050. Se os valores de 634 e 784 são substituídos para $P(t)$ e $P(t+\Delta t)$ nas fórmulas (7.1.a) e (7.1.b), com $\Delta t=35$, o resultado para a taxa anual é 0,006086 (0,6086%) e para a taxa instantânea 0,006067.

A Tabela 7.1 mostra as taxas anuais de crescimento de alguns países selecionados. Notam-se as taxas ainda muito elevadas de alguns países africanos, como Angola e Guiné Equatorial, e as taxas negativas do Japão, Rússia e Portugal. Também deve ser notado que países como o Brasil e a China, que pelo nível baixo do número médio de filhos das famílias já mostram uma predisposição para o declínio populacional, ainda aparecem com taxas positivas de crescimento. Este é um fenômeno transitório conhecido com *inércia demográfica*, ou seja, a tendência de países que tiveram padrões reprodutíveis tendentes a altas taxas de crescimento no passado, mas não mais na atualidade, a manter parte deste crescimento por algumas décadas depois da mudança no padrão reprodutivo, devido ao fato de que ainda continuam com uma estrutura etária jovem, tendo muitas mulheres em idade reprodutiva e poucos idosos. Por esta razão, além da influência da migração internacional em países como Cabo Verde e Macau, a taxa de crescimento nem sempre é um bom indicador do nível de reprodução de uma população. Este último conceito é medido de forma mais correta por uma outra taxa de crescimento, chamada a *taxa de crescimento intrínseco*, que será introduzida no Capítulo 22.

Tabela 7.1: Populações de países selecionadas em 2000 e 2020 (em milhares), com as respectivas taxas anuais de crescimento do período (percentuais)

País	População 2000	População 2020	Taxa Anual de Crescimento
Alemanha	81.401	83.784	0,14
Angola	16.395	32.866	3,54
Argentina	36.871	45.196	1,02
Bolívia	8.418	11.673	1,65
Brasil	174.790	212.559	0,98
Cabo Verde	428	556	1,31
Chile	15.342	19.116	1,11
China	1.290.551	1.439.324	0,55
Cuba	11.126	11.327	0,09
EUA	281.711	331.003	0,81
Federação Russa	146.405	145.934	- 0,02
França	59.015	65.274	0,51
Guiné-Bissau	1.201	1.968	2,50
Guiné Equatorial	606	1.403	4,29
Índia	1.056.576	1.380.004	1,34
Japão	127.524	126.476	- 0,04
Macau	428	649	2,11
México	98.900	128.933	1,33
Moçambique	17.712	31.255	2,88
Portugal	10.297	10.197	- 0,05
São Tomé & Príncipe	142	219	2,18
Timor-Leste	884	1.318	2,02

Fonte: Divisão de População das Nações Unidas, Revisão de 2019.

Para ajudar a visualizar as implicações de uma determinada taxa de crescimento a mais longo prazo às vezes se calcula o chamado *tempo de duplicação*, ou seja, o tempo necessário para que uma população que cresce com certa taxa alcance o dobro do seu tamanho inicial. Como existem efeitos de acumulação (crescimento sobre crescimento), o tempo que uma população que cresce a 2% por ano leva para duplicar não é 50 anos, mas só 35, e uma população que cresce a 4% por ano duplica em apenas 17,7 anos. As respectivas fórmulas são as seguintes:

$$Tempo\ de\ Duplicação = log(2) / log(1+r_{anual}) \approx ln(2) / r_{inst} = 0{,}6931 / r_{inst} \tag{7.4}$$

onde o log pode ter qualquer base, mas geralmente 10, e o *ln* por definição é o logaritmo de base *e*. Usando qualquer uma das taxas r_{anual} ou r_{inst} é fácil ver que, se o ritmo atual (entre 2015 e 2020) de crescimento da população latino-americana persistir, levará 73,7 anos para que ela duplique, um tempo bastante longo quando comparado com as tendências do passado.

7.3 A FÓRMULA BÁSICA DA CONTABILIDADE DEMOGRÁFICA E O CONCEITO DE COORTE

As populações podem ser descritas em termos de variáveis de estoque como de fluxo, mas estes dois aspectos não são independentes. As mudanças que ocorrem tanto no tamanho como na composição da população ao longo do tempo devem ser consistentes com os processos de mudança aos quais a população está exposta, descritos pelas variáveis de fluxo. Existem várias fórmulas de consistência que descrevem as relações que precisam ser satisfeitas. A mais conhecida e mais simples, que se aplica ao conjunto da população, é a *Fórmula Básica da Contabilidade Demográfica*, também conhecida por vários outros nomes como *Equação de Equilíbrio/Balanço Demográfico/Populacional*, *Equação Compensadora* ou *Equação de Concordância* ("Growth Balance Equation", em inglês), que simplesmente afirma que a diferença entre o tamanho total de uma população em dois momentos diferentes deve ser igual ao número de nascimentos ocorridos durante o período intermediário, menos o número de óbitos, mais o número de imigrantes e menos o número de emigrantes. Em termos mais formais:

$$P(t+n) = P(t) + N(t,t+n) - D(t,t+n) + I(t,t+n) - E(t,t+n) \qquad (7.5)$$

onde $P(t)$ e $P(t+n)$ são as populações existentes nos momentos t e $t+n$, e $N(t,t+n)$, $D(t,t+n)$, $I(t,t+n)$ e $E(t,t+n)$ se referem, respectivamente, aos nascimentos, óbitos, imigração e emigração ocorridos entre t e $t+n$. Qualquer divergência desta fórmula pode indicar uma de duas coisas:

- Uma ou mais quantidades que constam da fórmula podem ter sido medidas incorretamente; ou
- As unidades territoriais correspondentes às populações em t e $t+n$ podem não ser as mesmas, por exemplo porque se trata de um município ou uma província cujos limites foram modificados durante o período.

Populações que não têm migração ($I(t,t+n) - E(t,t+n) = 0$) são conhecidas como populações *fechadas*. A componente de crescimento da população que não envolve migração (ou seja, nascimentos menos óbitos) é chamada *vegetativo* (em inglês, "natural growth"). A terminologia "crescimento natural" às vezes é usada em português também, mas o termo "vegetativo" merece preferência.

A fórmula (7.5) é usada frequentemente para estimar $E(t,t+n)$ ou $I(t,t+n)-E(t,t+n)$ em circunstâncias onde todas as componentes da equação são conhecidas, com a exceção da migração. Para mais sobre este método residual para estimar a migração interna ou internacional, veja o Capítulo 11.

Mas não é só a evolução da população que precisa cumprir com certos requisitos de consistência. O mesmo acontece também com várias subpopulações. Por exemplo, o número de mulheres solteiras de 20-24 anos em 2010 foi diferente do número de solteiras de 25-29 anos em 2015, mas estes dois números têm algo em comum: a grande maioria das mulheres que fizeram parte do segundo grupo também já fazia parte do primeiro 5 anos antes. São as mulheres nascidas entre 1985 e 1989 cujo número de solteiras vai diminuindo no tempo na medida em que elas se casam ou unem e algumas morrem ou migram. Estas mulheres compõem uma geração ou, em linguagem propriamente demográfica, *uma coorte*. Uma coorte é um grupo de pessoas que passaram por um mesmo *evento* demográfico durante o mesmo período. Neste caso, o evento é o nascimento e o período 1985-1989. Uma coorte de nascimentos é o mesmo daquilo que comumente se chama uma geração. Mas a palavra "coorte" pode ser usada também para descrever, por exemplo, o conjunto de pessoas que se graduaram da escola secundária em 2008 ou o conjunto de mulheres que tiveram o seu primeiro filho entre 2010 e 2014.

É fácil ver que as duas quantidades mencionadas no parágrafo anterior se relacionam da seguinte forma:

- Mulheres solteiras de 25-29 anos em 2015 = Mulheres solteiras de 20-24 anos em 2010
- MAIS Mulheres pertencentes à coorte nascida entre 1985 e 1990 que entraram no país entre 2010 e 2015 enquanto ainda eram solteiras
- MENOS Mulheres pertencente à coorte nascida entre 1985 e 1990 que saíram do país entre 2010 e 2015 enquanto ainda eram solteiras
- MENOS Mulheres pertencentes à coorte nascida entre 1985 e 1990 que morreram entre 2010 e 2015 enquanto ainda eram solteiras
- MENOS Mulheres pertencentes à coorte nascida entre 1985 e 1990 que se casaram entre 2010 e 2015

Estas são todas as alternativas possíveis. Se esta conta não fechar, deve ser por causa de algum erro numa das componentes ou porque mudaram as fronteiras do país. Como no caso da Fórmula Básica acima, contas deste tipo são usadas com frequência na demografia para verificar a consistência das informações e principalmente para inferir o valor de uma das componentes, que talvez não possa ser medida diretamente, por meio das demais.

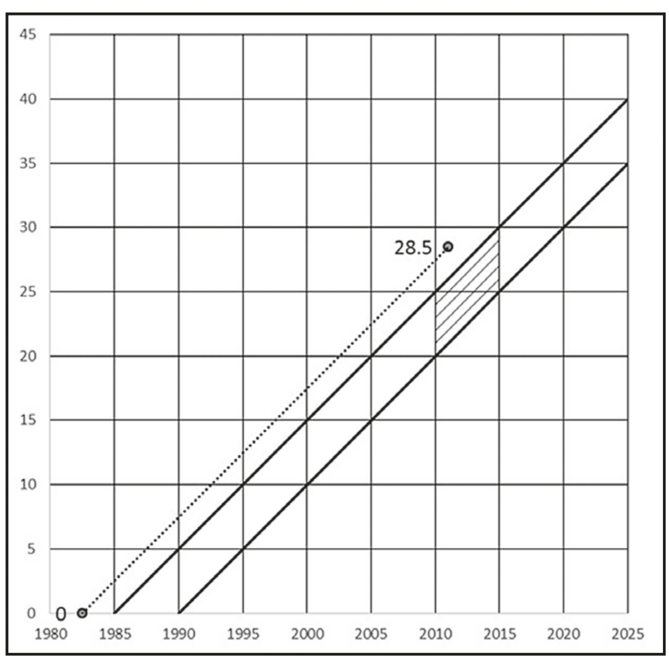
É de notar que as relações de consistência deste tipo se aplicam a coortes, mas não a grupos etários. Por exemplo, a priori não há nenhuma relação direta entre as mulheres solteiras de 20-24 anos em 2010 e em 2015, já as duas subpopulações pertencem a coortes diferentes (nascidas em 1985-89 e 1990-94, respectivamente). É de esperar que os dois números sejam mais ou menos semelhantes porque normalmente não ocorrem mudanças muito bruscas nos processos que determinam estes números, mas em teoria é inteiramente possível que sejam bastante diferentes devido a uma mudança rápida da natalidade por volta de 1990 ou a um mudança drástica no padrão de casamento de uma coorte para outra. Entretanto, como se verá no Capítulo 22, em populações onde os processos de mortalidade e fecundidade obedecem a certas condições de regularidade as relações entre um mesmo grupo etário em diferentes momentos do tempo sim acabam sendo mais previsíveis.

7.4 O DIAGRAMA DE LEXIS

A situação de um indivíduo ou de um grupo de indivíduos dentro da evolução demográfica de um país ou região pode ser caracterizada em termos de três características: 1. Tempo ou ano, mês e dia do calendário; 2. Idade exata ou intervalo de idade; e 3. Coorte ou data de nascimento. Estas características são dependentes entre elas: sabendo duas delas pode-se calcular a terceira. Por exemplo, as pessoas que compõem a coorte de nascimentos de 1995-1999 no 1º de janeiro de 2008 tinham 13-17 anos completos de idade. A situação se complica um pouco se ambas as características determinantes são dadas em termos de intervalos e não de números exatos. Por exemplo, durante o período de 2010-2014, a coorte nascida em 1985-1989 tinha entre 20 e 30 anos de idade. Mas não todas as pessoas que em algum momento deste período tiveram entre 20 e 30 anos de idade pertenciam à coorte de nascimentos de 1985-1989. Por exemplo, uma pessoa que no 1º de janeiro de 2011 tinha 28 anos nasceu em 1982 e, portanto, pertencia a uma coorte de nascimentos diferente.

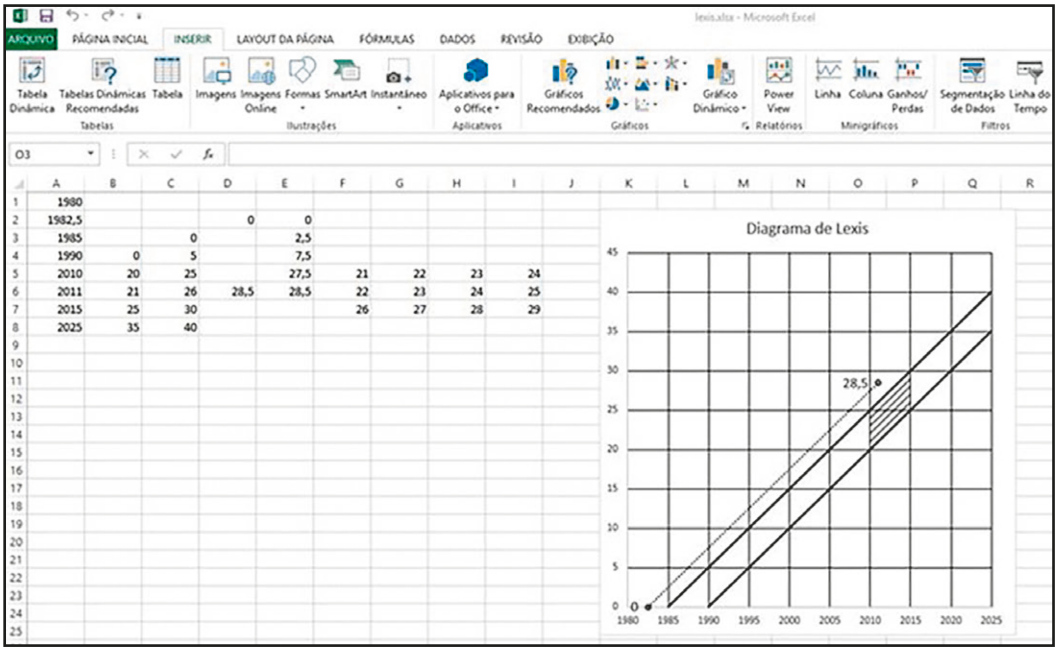
Para visualizar melhor as diferentes relações entre tempo, idade e coorte no que toca a características e fluxos de população usa-se na demografia um recurso chamado o *diagrama de Lexis*. O eixo horizontal do diagrama representa o tempo. O eixo vertical representa a idade. Cada vez que passa um ano calendário no tempo, uma pessoa fica um ano mais velha. Portanto, a trajetória da vida ou *linha vital* de cada pessoa é uma linha diagonal ascendente. A representação de uma coorte é um conjunto de linhas vitais que formam uma banda diagonal no diagrama. O Gráfico 7.1 mostra o diagrama de Lexis que representa os dados do parágrafo anterior. O paralelogramo com as linhas paralelas representa a coorte de nascimentos de 1985-1989 durante o período de 2010-2014 e o ponto marcado com "28.5" representa um indivíduo nascido no 1º de julho de 1982 que no 1º de janeiro de 2011 tem 28,5 anos de idade.

Gráfico 7.1: Exemplo de um diagrama de Lexis representando a coorte de nascimentos de 1985-1989 durante o período de 2010-2014 e um indivíduo de 28,5 anos no 1º de janeiro de 2011 com a sua linha vital desde o nascimento



Embora haja outras maneiras para fazê-lo, este diagrama foi montado em EXCEL usando as seguintes séries de dados:

Figura 7.1: Imagem de tela (ecrã) da composição do Gráfico 7.1 em EXCEL

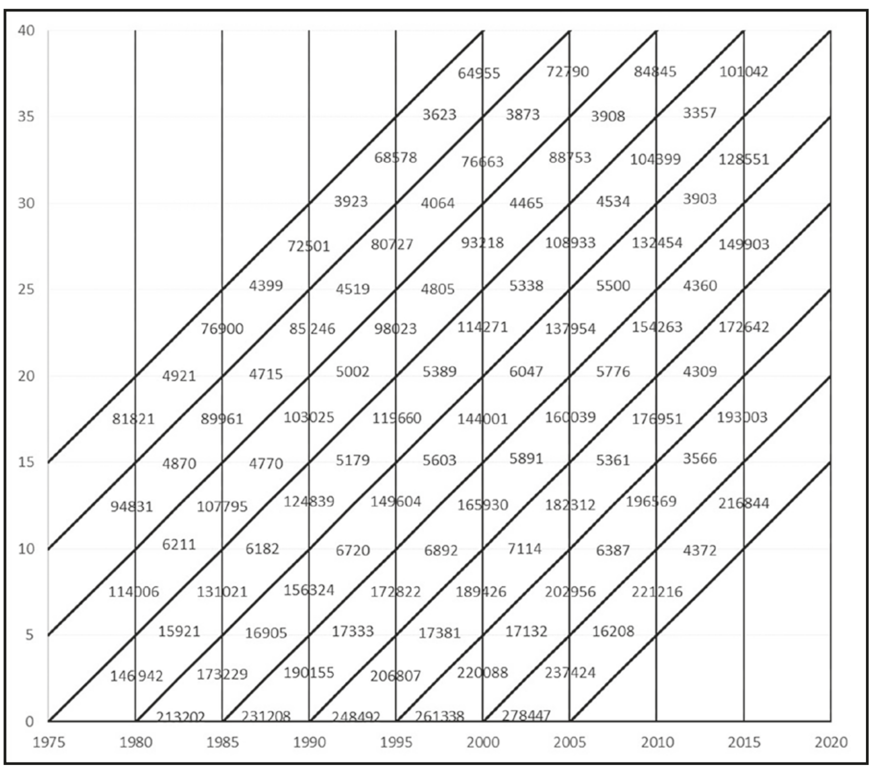


Todos os elementos do diagrama são Gráficos de Dispersão que têm a coluna A como argumento (Série X), sendo que a coluna D é representada como um Gráfico de Dispersão comum e as demais colunas como Gráficos de Dispersão com Linhas Retas. Para poder acomodar estas duas variedades dentro do mesmo diagrama, é preciso escolher a opção "Combo" na hora de definir o tipo de gráfico. As colunas B e C definem a coorte principal, D representa os dois pontos, E representa a linha diagonal pontilhada e F-I definem o padrão de linhas diagonais no intervalo 2010-2014. Os rótulos "0" e "28.5" só podem ser incluídos nas versões de EXCEL a partir de 2013 que possibilitam a edição de rótulos.

O problema básico da análise demográfica consiste na circunstância de que é extremamente difícil monitorar os três processos básicos (nascimentos, óbitos e migrações) de mudança simultaneamente em tempo contínuo. Embora seja perfeitamente legítimo representar a evolução de uma população desta maneira, não é a forma como o diagrama de Lexis normalmente é usado. Na prática, a maioria das fontes de dados não permite conhecer e desenhar todas as linhas vitais individualmente. Publicar a informação num formato que permitisse isso traria problemas de confidencialidade, além de gerar uma quantidade de detalhe que para a grande maioria dos usuários das estatísticas seria pouco funcional. Mesmo os Registros Civis mais sofisticados não publicam dados diários sobre todos os eventos, mas os agregam em intervalos de 1 ano calendário e intervalos etários de 1 ou 5 anos. Por outro lado, as características da população geralmente não são dadas num intervalo de tempo como o período 2010-2014 no Gráfico 7.1, mas como uma sequência de momentos discretos, por exemplo o 1º de janeiro de 2000, de 2005, de 2010 e de 2015. Nos espaços entre as linhas verticais (momentos no tempo) ou horizontais (idades exatas) são colocadas as quantidades de fluxo que levam às mudanças das características retratadas nas linhas. Neste contexto, é suficiente saber quantas linhas começam ou terminam numa determinada área e quantas cruzam determinadas barreiras como o limite entre dois anos (linha vertical) ou entre duas idades (linha horizontal).

O Gráfico 7.2 mostra um exemplo deste tipo de uso do diagrama, ilustrado com dados da Guiné-Bissau no período de 1980 a 2015. É importante assinalar que se trata de estimativas e interpolações baseadas em diferentes fontes e não em números diretamente observados nos censos. Aqui as linhas são verticais, indicando o 1º de janeiro do primeiro ano de cada quinquênio e os números de eventos se referem aos paralelogramos (losangos) entre quinquênios sucessivos da mesma coorte. As quantidades de fluxo retratadas aqui são simplesmente mudanças nos efetivos de população, sem distinção entre óbitos e migrações. Por exemplo, havia 196.569 pessoas com idades entre 10 e 14 anos no 1º de janeiro de 2010, das quais sobraram 193.003 no 1º de janeiro de 2015 quando tiveram entre 15 e 19 anos de idade. As demais 3.566 morreram ou migraram entre a primeira data e a segunda[1]. Na parte inferior do diagrama, na idade exata 0, aparecem os nascimentos ocorridos nos períodos de 1980-84 (213.202), 1985-89 (231.208), 1990-94 (248.492), 1995-99 (261.338) e 2000-04 (278.447).

Gráfico 7.2: Diagrama de Lexis das coortes nascidas entre 1960 e 2004 durante o período de 1980 a 2014 até os 39 anos de idade na Guiné-Bissau



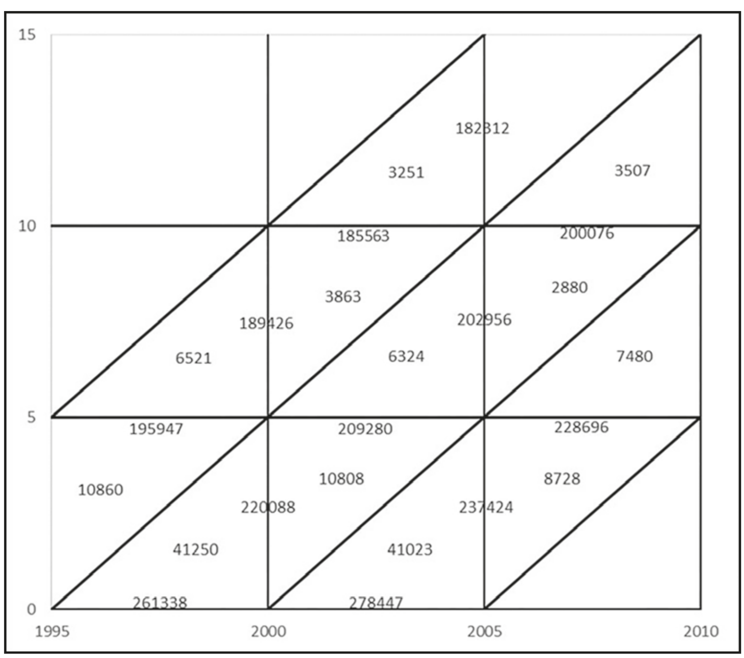
Fonte: Divisão de População da Nações Unidas, Revisão de 2015.

Seria possível construir um diagrama semelhante com linhas horizontais, indicando idades exatas e números de eventos em paralelogramos entre as idades exatas. Ou ainda seria possível (se os dados existem) fazer as duas coisas simultaneamente, com linhas horizontais para indicar

[1] Estritamente falando, as 193.003 pessoas de 15-19 anos presentes no 1º de janeiro de 2015 incluem *imigrantes* durante o período. A mistura de imigração com emigração representa um desafio analítico especial que será discutido em mais detalhe no Capítulo 11. Para não complicar o exemplo indevidamente, é melhor ignorar este detalhe por enquanto e assumir que os únicos eventos relevantes sejam a mortalidade e a emigração.

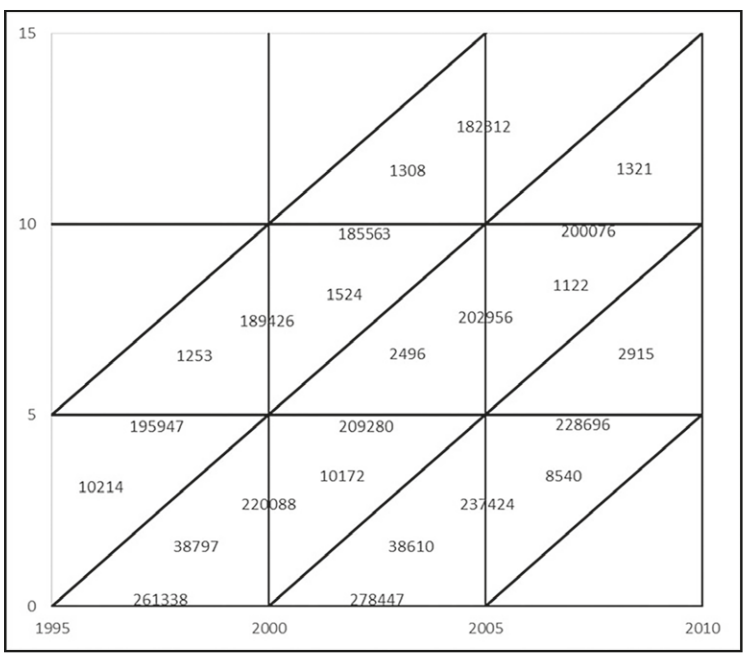
idades exatas, linhas verticais para indicar datas exatas e triângulos para indicar o número de eventos entre uma idade exata e uma data exata. O Gráfico 7.3 mostra uma parte do Gráfico 7.2 com dados hipotéticos para sugerir qual poderia ser a configuração de um diagrama deste tipo. Como o Gráfico 7.2, o diagrama mostra que 261.338 pessoas nasceram entre 1995 e 1999 e que destes 220.088 ainda estavam presentes no 1º de janeiro de 2000 e 202.956 no 1º de janeiro de 2005. Mas além disso, o Gráfico 7.3 mostra que 10.808 pessoas desta coorte morreram ou migraram antes do seu quinto aniversário, de modo que se celebraram 209.280 quintos aniversários. Os demais 6.324 morreram ou migraram antes do 1º de janeiro de 2005, mas já tendo mais de 5 anos. Para construir um diagrama como o Gráfico 7.3, em princípio é preciso dispor de um sistema de dupla classificação das pessoas em que os indivíduos são identificados tanto pela sua idade como pelo seu ano de nascimento. Poucos países publicam esta informação de forma sistemática e Guiné-Bissau não é um deles. Por isso, a construção das quantidades de fluxo nos triângulos geralmente passa pela aplicação de frações teóricas, para aproximar a divisão correta dos quadrados ou dos losangos em triângulos. Foi assim que se construiu o Gráfico 7.3.

Gráfico 7.3: Detalhe do Gráfico 7.2 em que se identificam simultaneamente os períodos quinquenais, intervalos etários e coortes



O exemplo nos Gráficos 7.2 e 7.3 é um pouco atípico no sentido de que a variável de fluxo retratada aqui nada mais é do que o aumento ou a diminuição da população das coortes entre duas datas ou duas idades exatas em função da mortalidade ou migração. Formulado desta maneira, o esquema cumpre com a equação de consistência para coortes. Mas normalmente este não é o caso porque a variável de fluxo geralmente é outra. Por exemplo, os números dentro dos triângulos poderiam referir-se só a óbitos e não a migrações. Esta situação é mostrada no Gráfico 7.4. Entretanto, os números nas linhas horizontais e verticais continuariam sendo influenciados pela migração de modo que a equação de consistência não se aplicaria mais. Os eventos retratados dentro dos triângulos poderiam, inclusive, ser de um tipo que não afeta o tamanho da população da coorte, como o número de anos completos que as crianças passam na escola. Uma criança que morre ou emigra sai da população e não consta mais dos efetivos de população na próxima data ou na próxima idade de referência. Mas isso não acontece com anos de escolaridade completada. Uma criança que completa um ano continua formando parte da população e depois pode completar um outro ano. Eventos deste tipo chamam-se *renováveis* e, como se verá na próxima seção, a sua interpretação no cálculo de taxas é um pouco diferente do tratamento de eventos *não renováveis* como óbitos. Em muitos casos é possível redefinir o processo de tal forma que se torne não renovável. Por exemplo, ter um filho é um evento renovável na vida de uma coorte de mulheres já que elas podem ter vários filhos durante um determinado período e continuam formando parte da coorte. Mas se o evento for definido como ter o *primeiro filho* e a coorte for de mulheres que ainda não passaram por esta experiência, o processo vira não renovável já que o nascimento do primeiro filho retira a mãe da coorte de mulheres nulíparas (que nunca tiveram filhos).

Gráfico 7.4: O mesmo diagrama do Gráfico 7.3, mas só com óbitos como eventos



Com a introdução do diagrama de Lexis dispõe-se agora dos elementos para quantificar a intensidade dos processos demográficos. Olhando o Gráfico 7.3, é preciso reconhecer que o número de eventos demográficos – sejam eles renováveis ou não renováveis – que ocorrem num determinado período a pessoas de uma determinada idade pode ser caracterizado de três maneiras distintas:

1. É possível analisar um determinado período (2000-2004) e um determinado grupo etário (5-9 anos) e somar os eventos *(3,863+6,324)*. Mas estes eventos pertencem a duas coortes diferentes, nascidas em 1990-1994 e 1995-1999, respectivamente.

2. É possível analisar um determinado período (2000-2004) e uma determinada coorte (os nascidos em 1995-1999) e somar os eventos assim *(10,808+6,324)*. Mas estes eventos caracterizam dois grupos etários diferentes, de 0-4 e de 5-9 anos, respectivamente.

3. A terceira possibilidade é analisar um determinado grupo etário (5-9) e uma determinada coorte (nascidos em 1995-1999) e somar os *6,324+2,880* eventos assim. Mas estes eventos se dividem entre dois períodos diferentes, de 2000-2004 e de 2005-2009, respectivamente.

Portanto, não há nenhuma maneira para analisar uma coorte única numa faixa etária única dentro de um período único. Entre as três alternativas possíveis, as mais usadas são a primeira, conhecida como *análise de período*, e a terceira, conhecida como *an*álise de coorte. A segunda é pouco usada, por misturar diferentes grupos etários. Análises que acompanham uma coorte também são chamadas *longitudinais*, especialmente quando fazem este acompanhamento durante vários períodos. Análises que misturam diferentes coortes são chamadas *transversais*, especialmente quando abrangem uma sequência de grupos etários.

Embora o princípio do diagrama de Lexis seja fácil de entender, a relação exata entre a visão de idade, período e coorte em realidade apresenta alguns desafios conceituais. Vandeschrick (1995), por exemplo insiste que o diagrama realmente trata de três dimensões, embora as retrate como apenas suas. Ao analisar a evolução dos processos demográficos faz todo o sentido distinguir entre efeitos de idade (ao envelhecer, o risco de morte aumenta), período (certos períodos históricos foram caracterizados por uma mortalidade mais alta) e coorte (certas coortes são menos resistentes porque estiveram expostas a eventos traumáticos na infância). Entretanto, essa análise em três componentes se confronta com o fato de que elas não são independentes, pois Período – Idade = Coorte (de Nascimento), o que impossibilita certos tipos de análise estatística como a regressão múltipla. O problema é conhecido como o problema de *identificação* e existem diferentes técnicas para lidar com ele (Yang e Land, 2006, 2013; Fu, 2018). A questão será brevemente abordada na seção 13.5 do Capítulo 13, mas o seu tratamento detalhado está além dos propósitos deste livro.

7.5 TAXAS E PROBABILIDADES

O último parágrafo da seção anterior discute as opções para a contagem do número de eventos relevantes para quantificar a intensidade de uma variável de fluxo. Mas para construir um indicador de intensidade, esta quantidade de eventos precisa ser relacionada com um denominador que de alguma forma representa o número *potencial* de eventos que poderiam ter acontecido. O conceito mais correto para expressar este número potencial é o *número de anos-pessoa* vividos pela população dentro da área relevante do diagrama de Lexis, ou seja, o comprimento conjunto, dentro da área, de todas as linhas vitais que cruzam a área de forma completa ou parcial. Esta ideia é mais fácil de entender no caso de uma coorte como a coorte de nascimentos de 1995-1999 no Gráfico 7.3 ou Gráfico 7.4. No losango que define a vivência desta coorte para as idades de 5-9 anos, inicialmente houve 209.280 pessoas, das quais no final sobraram 200.076. Cada uma destas pessoas contribuiu 5 anos ao conjunto de anos-pessoa vividos pela coorte, ou seja, um total de $5 \cdot 200.076 = 1.000.380$ anos.

Mas também houve *6.324+2.880 = 9.204* pessoas que não completaram o intervalo e cujas linhas vitais pararam em algum lugar intermediário, seja porque morreram ou porque emigraram. Muitas vezes se supõe que estas pessoas contribuíram a metade do período, ou seja 2,5 anos cada uma, de modo que o número total de anos-pessoa acaba sendo a média da população inicial e final vezes o tamanho do intervalo. Mas isso nem sempre é correto. Sabe-se, por exemplo, que as crianças que morrem no primeiro ano de vida tendem a morrer muito mais no início (o primeiro mês) do que mais tarde (ver Capítulo 8). Portanto, o comprimento médio das linhas vitais destas crianças seria bem menos do que a metade do período. Em teoria, seria possível calcular todas as contribuições feitas por todas as pessoas que saíram da população durante o período, mas na prática raramente dispõe-se de dados suficientemente detalhados para fazer isso. Quando não se pode supor que cada pessoa que não completa o intervalo contribui a metade do mesmo, a solução geralmente adotada na prática é a aplicação de *fatores de separação,* baseados na experiência, para quantificar a contribuição das linhas vitais incompletas. Neste caso o fator de separação para o intervalo de 5 a 9 anos provavelmente seria bem próximo a 0,5, talvez 0,48. Sendo assim, o número de anos-pessoa seria

$$Anos\text{-}pessoa = 5 \cdot 200.076 + 0,48 \cdot 5 \cdot 9.204 = 1.022.469,6\ anos \tag{7.6}$$

Como se percebe, o critério de anos-pessoa pode ser o mais correto do ponto de vista conceitual, mas na prática o seu cálculo exige certas aproximações. Isso já é o caso numa análise de coorte, mas ainda mais numa análise de período. A área relevante neste caso é um quadrado, por exemplo a faixa etária de 5-9 anos em 1995-1999, de modo que o resultado depende de onde exatamente as linhas vitais cruzam este quadrado, mais perto da diagonal (onde o seu comprimento seria maior) ou mais perto dos cantos (onde seria menor). Entretanto, na prática este tipo de detalhes normalmente não é levado em conta e o número de anos-pessoa é estimado de uma das três seguintes formas:

1. Como a média da população no início e no fim do período, ou seja, *5 · (189.426+202.956) / 2 = 980.955* anos.

2. Eventualmente, se existem razões para supor uma distribuição desequilibrada, pode-se usar uma ponderação usando fatores de separação, como *5 · (0,48 · 189.426 + 0,52 · 202.956).*

3. Se existe uma estimativa para a população na metade do período, esta pode ser usada também, multiplicada pelo tamanho do intervalo.

Os resultados de cada um destes procedimentos serão ligeiramente diferentes, mas normalmente as diferenças não devem ser muito significativas. Na prática a primeira e a terceira solução são as preferidas.

Como foi mencionado no Capítulo 6, taxas são razões, mas nem toda razão é uma taxa. A particularidade de uma taxa é que o seu numerador mede o número de eventos que ocorrem num determinado período enquanto o denominador se refere à população que pode ser o objeto deste evento[2]. Quando a população se limita a uma determinada faixa etária, distinguem-se dois tipos de taxas:

1. A taxa *central* (ou do tipo "m", na terminologia de Hinde, 1998) combina eventos segundo o critério de período no numerador com uma população exposta segundo o número de anos-pessoa ou, na prática, segundo o critério 1 ou 3 do parágrafo anterior.

2. A taxa *inicial* (ou do tipo "q", na terminologia de Hinde) combina eventos segundo o critério de coorte no numerador com a população presente no início da faixa etária.

Ambas as taxas têm vantagens e desvantagens. A vantagem da taxa central é que ela costuma ser fácil de calcular, principalmente se forem usadas as aproximações 1 ou 3 do parágrafo anterior e que ela se refere a um período único, enquanto a segunda taxa necessariamente combina informação de dois períodos. A desvantagem da taxa central é que ela mistura a experiência de coortes distintas. Por esta razão, ela também é chamada *transversal*. Ela não pode ser interpretada como uma probabilidade e nem como proporção. A taxa inicial é mais difícil de calcular, mas ela retrata a experiência de uma coorte real, ou seja, é uma medida *longitudinal*. No caso de eventos não renováveis, ela é uma proporção porque as pessoas afetadas pelos eventos no denominador fazem parte da população inicial no denominador. Esta proporção também pode ser interpretada como uma probabilidade.

Considere-se a o grupo etário de 5-9 anos e o período de 2000-2004 no Gráfico 7.4. O numerador da taxa central de mortalidade para este grupo é *1.524+2.496 = 4.020* óbitos. O denominador é mais facilmente calculado como *5·(189.426+202.956)/2 = 980.955*. Portanto, a taxa (multiplicada por 1.000) é

$$ {}_5m_5 = 1.000 \cdot 4.020 \,/\, 980.955 = 4{,}10 \text{ por } 1.000 \tag{7.7} $$

Por outro lado, a taxa inicial para os períodos 2000-2004 e 2004-2009 tem um numerador de *2.496+1.122 = 3.618* e um denominador de 209.280, de modo que

$$ {}_5q_5 = 1.000 \cdot 3.618 \,/\, 209.280 = 17{,}3 \text{ por } 1.000 \tag{7.8} $$

Como a mortalidade é um processo não renovável, este último número pode ser interpretado como a probabilidade de sobrevivência da idade exata de 5 anos até a idade exata de 10 anos no período entre 2000 e 2009. No caso de eventos renováveis, como nascimentos, o cálculo destas taxas é o mesmo, mas raramente se calculam taxas do tipo "q" para estes eventos e quando são calculadas elas não podem ser interpretadas como probabilidades. A sua interpretação correta é o número esperado de vezes que o evento ocorre a cada pessoa dentro da faixa etária. É de notar também que o denominador da taxa central depende da amplitude do período e o denominador da

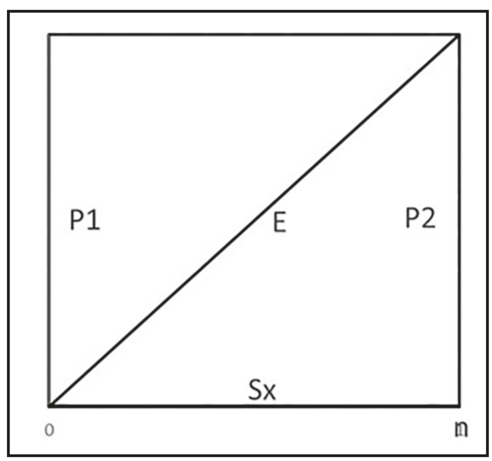
2 Estritamente falando, segundo esta definição uma taxa de crescimento como definida em (6.1.a) ou (6.1.b) não seria uma taxa, mas esta terminologia está tão enraizada que não vale a pena mudá-la.

taxa inicial não. Por isso (além do fato de que (7.7) e (7.8) não se baseiam exatamente nos mesmos dados) o resultado de (7.8) é quase 5 vezes maior do que (7.7).

7.6 CONVERSÃO DE TAXAS CENTRAIS EM TAXAS INICIAIS

Como foi notado acima, uma das diferenças fundamentais entre taxas centrais e iniciais é que as taxas centrais se aplicam a um único período enquanto as taxas iniciais necessariamente se estendem por dois períodos. Entretanto, é comum encontrar situações (ver Capítulos 8 e 9) onde só se dispõe de taxas centrais referentes a um período, mas onde é preciso estimar probabilidades. Isso exige algum tipo de mecanismo aproximado para converter taxas do tipo *m* em taxas do tipo *q*. O seguinte procedimento é uma maneira para desmembrar um quadrado do diagrama de Lexis num losango para estimar uma taxa inicial.

Gráfico 7.5: Quadrado com tamanho de n por n anos, com população P1 no ano inicial e P2 no ano final, E eventos não renováveis e Sx pessoas que completaram x anos de idade



Supõe-se que não há informação sobre a divisão dos eventos *E* entre os dois triângulos que compõem o quadrado com tamanho de *n* por *n* anos do Gráfico 7.5, ou seja, que não se dispõe de uma dupla classificação. O primeiro passo portanto consiste em dividir os eventos *E*. O recurso geralmente usado para este propósito são os fatores de separação já introduzidos acima. Supondo que o fator *a* simboliza a proporção dos eventos *E* que pertencem ao triângulo superior da esquerda, logicamente o número de eventos nesse triângulo seria *a·E* e no outro triângulo *(1-a)·E*.

O próximo passo consiste em deslocar o triângulo da esquerda para a direita, para formar um losango. Mas isso não pode ser feito diretamente porque a população *P1* pode ser diferente de *P2*. Portanto, todas as quantidades do triângulo superior da esquerda são multiplicadas por *P2/P1*. Assim, *P1* se transforma em *P2* e *a·E* em *a·E·P2/P1*. Portanto o número ajustado de eventos é *E·(a·P2/P1+1-a)*. Este número de eventos deve ser dividido pelo número inicial de pessoas entrando na faixa etária, que é S_x. Portanto,

$$ {}_nq_x = \frac{E\,\left(a\frac{P2}{P1} + 1 - a\right)}{Sx} \qquad (7.9)$$

Substituindo a definição da taxa central ${}_nM_x$, a seguinte fórmula aparece:

$$ {}_nq_x = n\ {}_nM_x\ \frac{(P1+P2)\big((1-a)\ P1+a\ P2\big)}{2\ Sx\ P1} \tag{7.10} $$

Esta fórmula ainda depende de várias incógnitas. Além da incerteza sobre o fator de separação *a* também há o fator *Sx*, cujo valor geralmente não se conhece exatamente. No contexto da tábua de vida, que será discutida no Capítulo 9, existem métodos padronizados para preencher essas lacunas. Aqui basta assinalar que em situações onde o evento *E* é responsável por (quase) toda a diferença entre *Sx* e *P2*, é razoável supor que *Sx* = *P2* + ½ E. Se além disso não houver muita diferença entre *P1* e *P2*, a fórmula (7.10) pode ser aproximada da seguinte maneira:

$$ {}_nq_x \approx \frac{n\ {}_nM_x}{1+\frac{1}{2}n\ {}_nM_x} \tag{7.11} $$

Esta fórmula constitui uma aproximação razoável em casos onde a modificação dos efetivos de população se deve basicamente a um único processo não renovável, como a mortalidade. Mas em casos como o retratado nos Gráficos 7.3 e 7.4, onde existe uma interferência significativa da migração, os resultados são menos próximos. Isso pode ser ilustrado com o período 2000-2004 e a faixa etária de 5-9 anos no Gráfico 7.4. Como neste caso existe uma dupla classificação do evento *E*, não é preciso supor a homogeneidade da mortalidade dentro de todo o quadrado e o mesmo pode ser desmembrado sem supostos adicionais. Isso dá como resultado uma taxa inicial de 19,73 por 1.000. Se em vez disso for usada a fórmula (7.9), que supõe a homogeneidade da mortalidade, o resultado obtido é uma taxa de 19,86 por 1.000. Mas se for usada a fórmula (7.10) substituindo o resultado de (7.7) para ${}_nM_x$, a estimativa acaba sendo 20,29 por 1.000. Este exemplo mostra que é preciso ter certo cuidado na aplicação de (7.11) em circunstâncias onde o processo estudado não é puro, mas sofre a influência de outros eventos. Por outro lado, se a incidência destes outros eventos for conhecida explicitamente, pode-se escrever *Sx* = *P2* + ½ *E1* + ½ *E2* e elaborar uma fórmula alternativa para (7.10) que contempla tanto *E1* como *E2* no denominador.

7.7 COORTES SINTÉTICAS

Em muitos sentidos a forma ideal de observação do comportamento demográfico é o acompanhamento de uma mesma coorte desde o evento que a criou até a última ocorrência do fenômeno estudado, por exemplo desde o nascimento até a morte do último sobrevivente. Ao fazer isso, a esperança de vida pode ser definida simplesmente como a idade média com que os integrantes da coorte morreram. Entretanto, este processo de observação pode demorar muito tempo (no caso, mais de 100 anos) e no final a informação assim recolhida, por mais detalhada que seja, pode já não ter muita relevância prática. Fora algumas aplicações muito específicas, qual é a utilidade prática de uma descrição muito detalhada da história de sobrevivência e morte das pessoas nascidas há mais de 100 anos ? Na maioria dos casos haverá muito mais interesse nos eventos demográficos mais recentes experimentados por diferentes coortes da população. Em vez de medir qual foi a média de anos vividos pela coorte nascida há mais de 100 anos, é muito mais interessante quantificar quantos anos as pessoas vivem em média hoje em dia. Para poder responder esta pergunta usa-se um recurso conhecido como "coorte sintética".

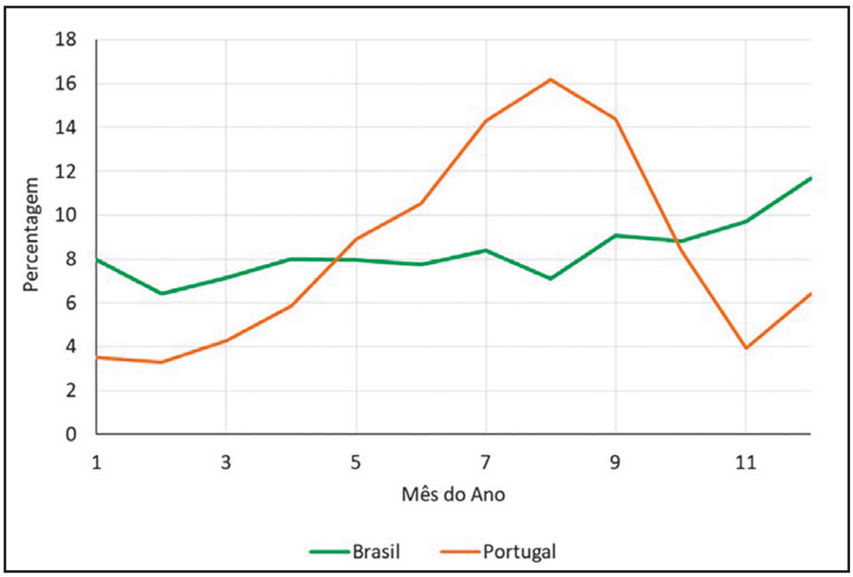
A coorte sintética se constrói em duas etapas. A primeira é a conversão de taxas centrais em taxas iniciais, de acordo com o explicado na seção anterior. A segunda etapa consiste em concatenar estas taxas numa sequência de idades ou intervalos etários sucessivos, como se representassem a vida de uma única coorte. Noutras palavras, se apresenta uma coluna de taxas transversais, representando coortes diferentes, como se fosse a história longitudinal de uma só coorte. O resultado é um tipo de "coorte Frankenstein" em que a experiência atual da coorte de 0-4 anos (nascida há 0-5 anos) define a história de vida das pessoas na faixa dos 0-4 anos, a experiência atual da coorte de 5-9 anos (nascida há 5-10 anos) define a história de vida das pessoas na faixa dos 5-9 anos, e assim por diante. Este procedimento permite construir uma sequência de eventos e taxas e calcular indicadores como a esperança de vida ou o número médio de filhos tidos das mulheres, de uma forma que descreve a atualidade e não uma sequência de tendências do passado. Entretanto, é importante frisar que a coorte na qual esses cálculos se baseiam não representa a experiência de vida real de qualquer grupo de pessoas. Ou seja, é uma ficção estatística. Quando se diz que "a esperança de vida na Guiné-Bissau em 2010-2014 era 54,7 anos", isso não significa (como às vezes se afirma) que "uma criança nascida na Guiné-Bissau em 2010-2015 na média viverá 54,7 anos", pois não há como saber isso. O que significa é: "Se uma criança nascida na Guiné-Bissau em 2010-2014 ao longo da sua vida fosse exposta às condições de mortalidade que naquele período existiam no país em cada idade, em média viveria 54,7 anos".

Na medida em que o comportamento demográfico das pessoas for determinado puramente pelas circunstâncias atuais, sem "memória demográfica" (o que os estatísticos chamam um "processo de Markov"), as coortes sintéticas representam bastante bem o que acontece nas coortes reais que as compõem. Mas quanto mais relação houver entre o comportamento das pessoas em diferentes fases das suas vidas, mais enganosas podem ser as estatísticas derivadas de coortes sintéticas. Este tipo de problema pode existir no caso da mortalidade, por exemplo se uma determinada epidemia que afetou as pessoas como crianças deixou certas vulnerabilidades no combate a doenças mais tarde na vida. Mas é na análise do comportamento reprodutivo que as limitações do conceito de coorte sintética ficam mais evidentes. Suponha-se, por exemplo, – hipoteticamente – que por algum motivo todas as mulheres de todas as idades decidam adiar a sua próxima gravidez em 1 ano (supondo que tenham os meios para isso). O resultado seria que o número de filhos nascidos na coorte sintética atual cairia para zero, enquanto no próximo ano aumentaria para mais ou menos o dobro do normal. Mas este comportamento errático das taxas nas coortes sintéticas não afetaria as coortes reais que manteriam mais ou menos o mesmo número de filhos ao longo da história de vida das mulheres, com apenas uma flutuação modesta no "timing" dos nascimentos. Este tipo de problemas e algumas soluções para lidar com eles será discutido no Capítulo 10.

7.8 A SAZONALIDADE DOS EVENTOS DEMOGRÁFICOS

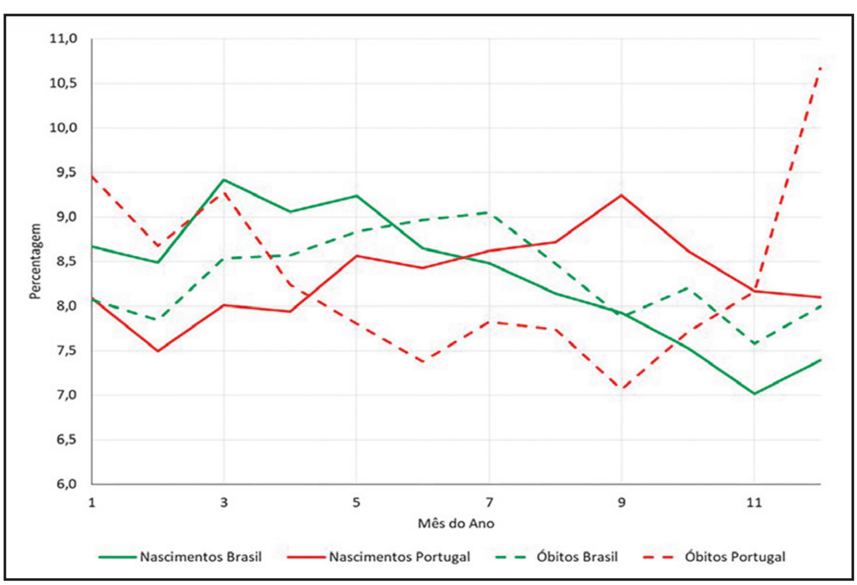
Diferentemente dos indicadores econômicos, muitos dos quais têm uma periodicidade trimestral ou até mensal, os indicadores demográficos geralmente são calculados para períodos mais longos, o que reduz a necessidade de cuidados com flutuações sazonais. Entretanto, ocasionalmente se fazem comparações de mais curto prazo que precisam de alguns cuidados neste sentido. Por exemplo, um artigo de jornal poderia notar que o número de casamentos registrados no mês de maio de um determinado ano foi consideravelmente menor do que 6 meses antes e poderia interpretar isso como um sinal conjuntural de diminuição na confiança no futuro. Mas isso não toma em conta que, apesar da fama do mês de maio de ser o "mês das noivas", os meses com as mais altas incidências de casamentos no Brasil são justamente novembro e dezembro. Portanto, seria mais prudente fazer a comparação com o mês de maio do ano anterior, como se faz também com indicadores econômicos afetados pela sazonalidade.

Gráfico 7.6: Distribuição mensal de casamentos ocorridos no Brasil e em Portugal em 2016 (por cem)



Fontes: Estatísticas do Registro Civil do Brasil e de Portugal.

Gráfico 7.7: Distribuição mensal de nascimentos e óbitos ocorridos no Brasil e em Portugal em 2016 (por cem)



Fontes: Estatísticas do Registro Civil do Brasil e de Portugal.

O Gráfico 7.6 mostra que a incidência de casamentos no Brasil é mais elevada no final da primavera e no início do verão, mas a variação não é muito forte. Em Portugal, a variação é muito maior e os casamentos se concentram no final do verão do hemisfério norte (julho-setembro). Os nascimentos também são afetados por certo grau de sazonalidade. No Brasil a sazonalidade dos casamentos e dos nascimentos tem magnitudes comparáveis, com o maior número de nascimentos em março, abril e maio e o menor número nos últimos meses do ano (Gráfico 7.7). Em Portugal a sazonalidade dos nascimentos é menor do que dos casamentos. Apesar disso, Caleiro (2008) mostra que há uma relação causal entre ambos, mesmo num país onde a fecundidade é baixa e maioritariamente planejada. Para maiores detalhes sobre a sazonalidade dos nascimentos, ver Moreira (2008). Talvez o mais surpreendente seja que existe também uma sazonalidade nos óbitos, com um claro aumento no inverno (maio-julho no Brasil, dezembro-março em Portugal). Os meses de menor incidência de mortes no Brasil são novembro-fevereiro).