

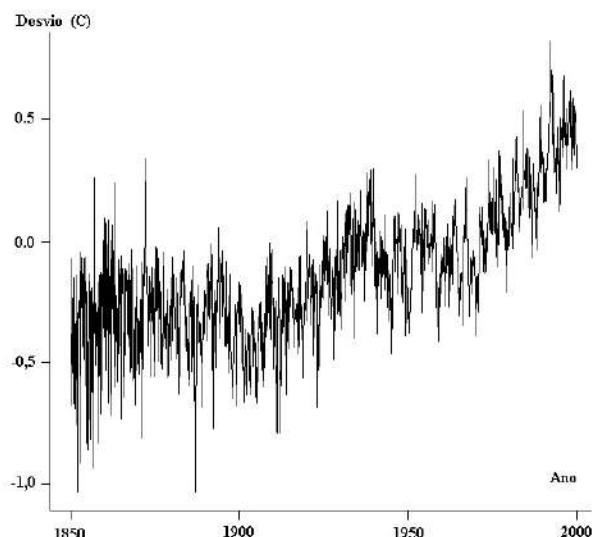
## Capítulo 2. Conceitos e ferramentas básicas

- 2.1. Gráficos
  - 2.2. Componentes de uma série temporal:
    - 2.2.1. Introdução
    - 2.2.2. Nível e tendência / ciclo
    - 2.2.3. Sazonalidade
    - 2.2.4. Resíduo ou erro aleatório
    - 2.2.5. Tipos especiais de séries temporais
  - 2.3. Medidas de erro de previsão
    - 2.3.1. Terminologia e notação, erros de previsão
    - 2.3.2. Acurácia e precisão
    - 2.3.3. Medidas de erro de previsão baseadas em médias aritméticas
    - 2.3.4. Método ingênuo de previsão (método *naïve*)
  - 2.4. Tipos de métodos para o estudo de séries temporais
    - 2.4.1. Métodos para previsão
    - 2.4.2. Métodos para análise
- 

Neste capítulo, veremos alguns dos conceitos e ferramentas básicas que devem ser conhecidos antes que começemos a estudar com detalhes cada família de métodos de previsão ou análise. Na seção 2.1, veremos os gráficos mais usados; na seção 2.2, os componentes nos quais uma série pode ser analisada, e alguns tipos especiais de séries em que esta análise não é possível; na seção 2.3, as medidas do erro de previsão; na seção 2.4, uma breve visão geral dos métodos existentes para previsão e análise.

### 2.1. Gráficos

A ferramenta básica para o estudo de séries temporais é o “gráfico de linha” (*line-plot* ou *timeplot*), que mostra a evolução de uma série ao longo do tempo. A variável de interesse  $Z$  é representada no eixo vertical, e o tempo  $t$  no horizontal.



**Figura 1. Variação da temperatura média mundial, 1850-2005**  
série *globtp.txt* [1]

Entre todos os tipos de gráficos usados na Estatística, este é com certeza o mais freqüentemente encontrado na mídia não-especializada (jornais, revistas, internet). A Fig. 1 mostra um exemplo; nele, fica bem evidente a elevação progressiva da temperatura mundial, representada aqui pelos desvios desta temperatura em relação à média mensal. Outros exemplos, para séries de vários tipos, serão mostrados na seção 2.2.

Note que, na realidade, estes gráficos deveriam mostrar um conjunto de pontos separados entre si, já que em geral a variável de interesse não é contínua, e sim medida em intervalos discretos. O gráfico da Fig. 2A, por exemplo, mostra uma série de temperaturas médias mensais; o valor medido a cada mês é representado por um ponto. Para facilitar o entendimento, porém, o gráfico geralmente inclui uma linha conectando os pontos, como na Fig. 2B; isto ajuda a destacar a direção que a série está tomando a cada instante.

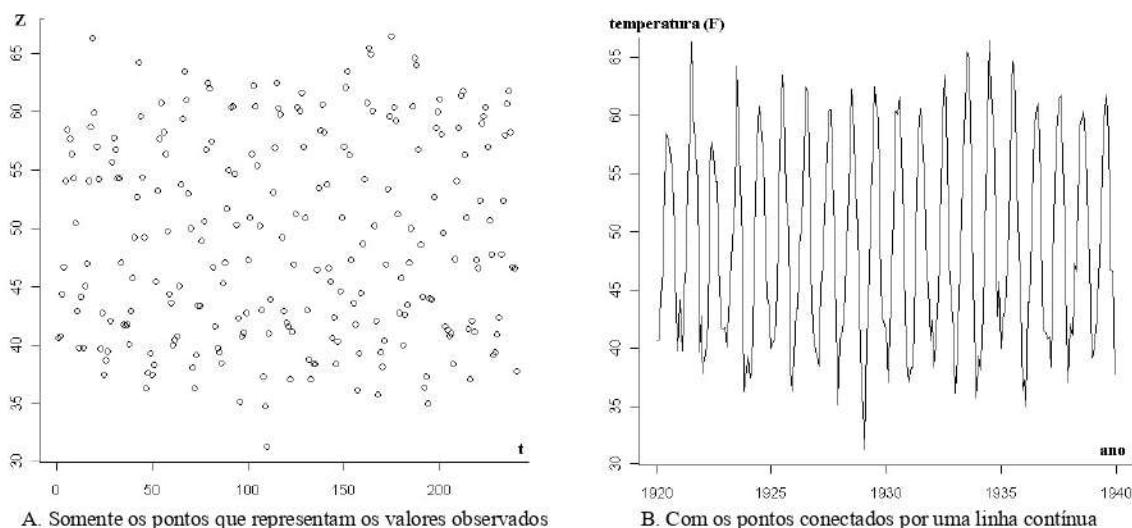
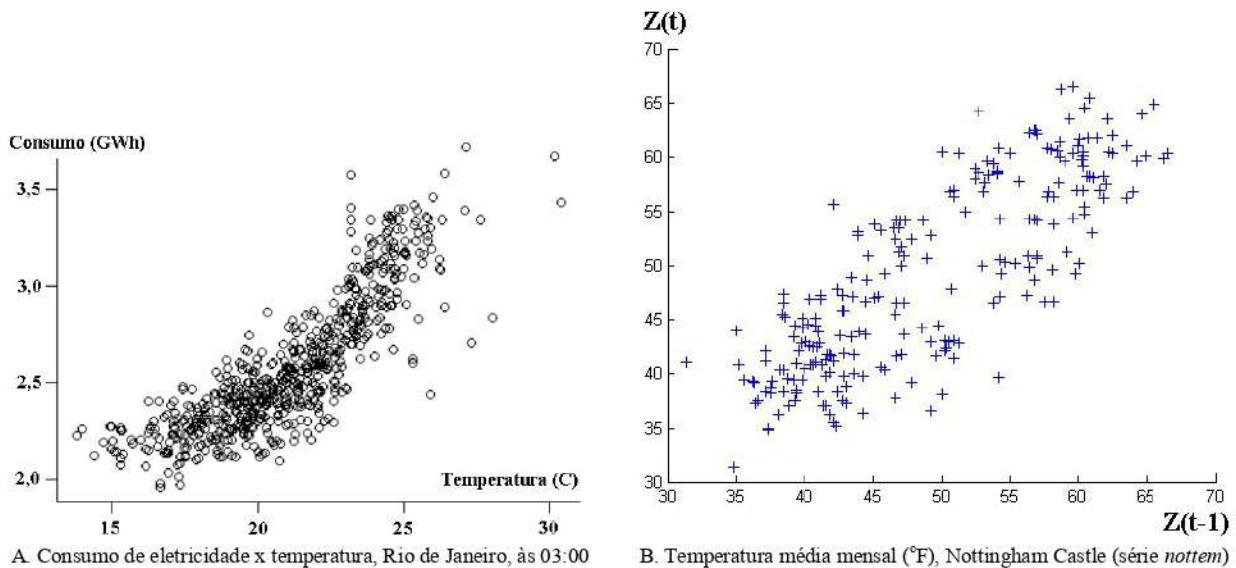


Figura 2. Temperatura média mensal ( $^{\circ}\text{F}$ ), em Nottingham Castle (série *nottem*)

Outro tipo de gráfico bastante usado em APST é o “diagrama de dispersão” (*scatterplot*), que ilustra o efeito que uma variável *independente* (no eixo horizontal) tem sobre uma variável *dependente* (no eixo vertical). Um exemplo deste gráfico já foi visto na Fig. 4 do Cap. 1, que mostra como o consumo de eletricidade de uma cidade é afetado pela temperatura do ar. A Fig. 3 ilustra a mesma relação consumo x temperatura, mas com dados do Rio de Janeiro, medidos às 03:00 horas da manhã de cada dia. Como o Rio é uma cidade quente, onde praticamente não existem aparelhos elétricos de aquecimento, o gráfico não tem agora a forma aproximada de um “U”; há apenas a parte direita deste U, correspondente às temperaturas mais elevadas.

O diagrama da dispersão também é bastante usado em APST para mostrar como o valor da variável em um instante depende (probabilisticamente) do valor desta mesma variável em instantes anteriores. O diagrama da Fig. 3B foi construído com a mesma série de temperaturas já usada na Fig. 2, e mostra que há uma relação entre os valores observados a cada instante - a temperatura média de cada mês depende da temperatura média do mês anterior. Esta dependência dos valores da série entre si é chamada de “autocorrelação”, e é nela que se baseiam vários modelos estatísticos de previsão (estes modelos formam uma família denominada *ARIMA*, e serão vistos a partir do Cap. 6).



**Figura 3. Exemplos de diagrama de dispersão**

Um terceiro tipo de gráfico usado muito freqüentemente é o “correlograma”, que mostra os valores da autocorrelação existentes em uma série, como função do desfase (intervalo de tempo) entre cada par de observações. Como este gráfico é tipicamente usado no trabalho com modelos ARIMA, iremos abordá-lo no Cap. 6.

## 2.2. Componentes de uma série temporal

### 2.2.1. Introdução

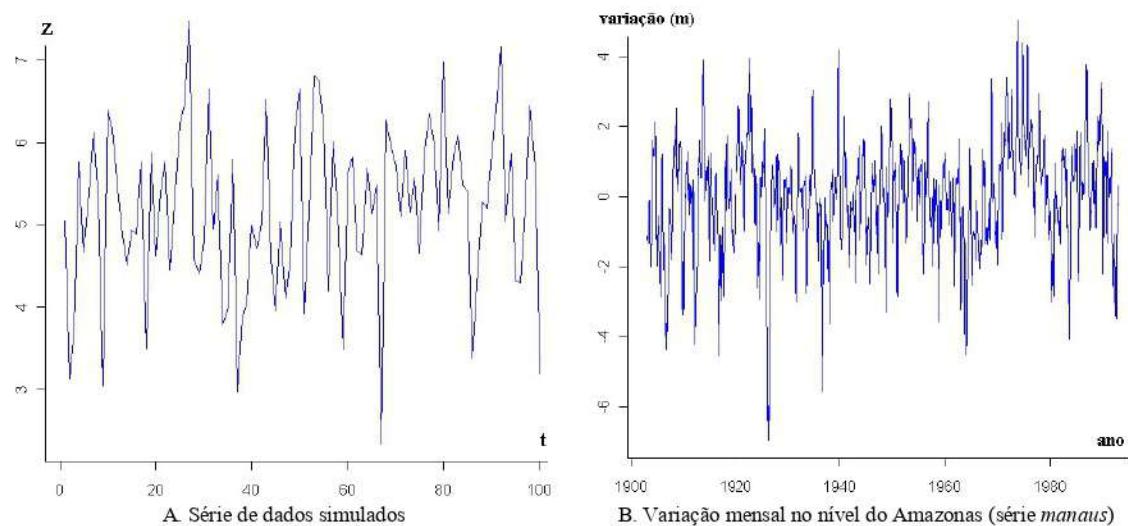
“Analisar” um objeto significa decompor este objeto em suas partes componentes, e mostrar como elas se relacionam. Na *análise sintática* tradicional, por exemplo, uma oração é decomposta em *sujeito, verbo, e predicado*, depois cada um destes componentes é decomposto em outros subcomponentes (predicado = verbo + objeto, etc.); na *análise química*, uma substância é decomposta nos elementos químicos que a compõem.

Nas séries temporais, geralmente a decomposição é feita em dois componentes básicos, um *padrão* e um *erro aleatório*. O padrão é a parte regular da série, que é previsível, pelo menos aproximadamente; se extrairmos da série o padrão, o que resta é o resíduo ou *erro aleatório*, imprevisível. O padrão pode às vezes ser decomposto em *nível, tendência, ciclo e sazonalidade*. Estes componentes devem ser interpretáveis, e fáceis de distinguir num gráfico; as séries encontradas na prática podem ser combinações destes componentes, em proporções diferentes. Neste capítulo, discutiremos como estes componentes podem ser revelados a partir de uma análise gráfica, baseada nas idéias dos métodos de *Decomposição Clássica* (Cap. 3). Estes componentes (nível, tendência, sazonalidade, resíduo) serão úteis quando estudarmos métodos de previsão nos Caps. 4 e 5. Os modelos ARIMA (Cap. 6 a 15) usam um conceito diferente para tendência.

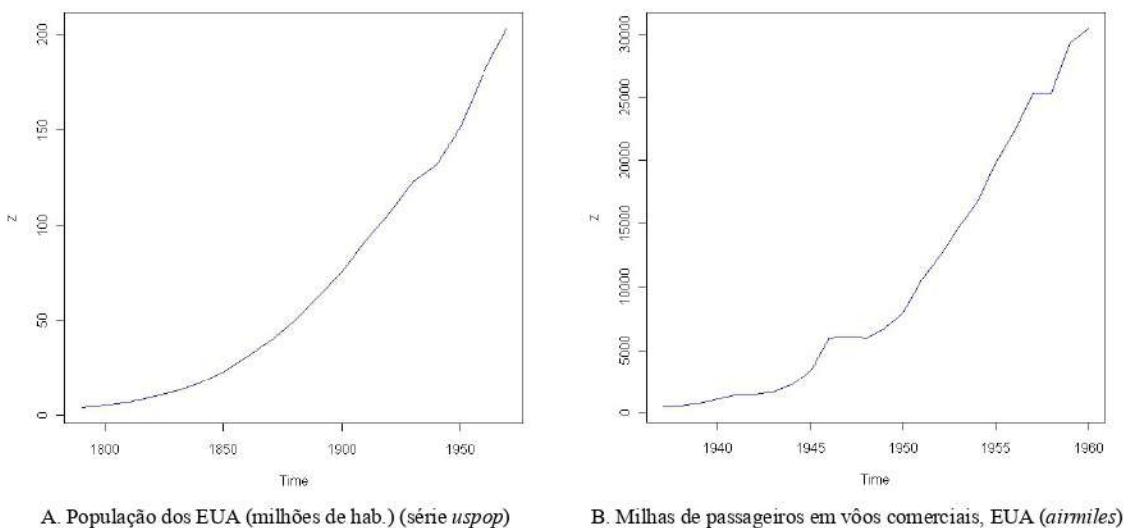
### 2.2.2. Nível e Tendência / Ciclo

Chamamos de *nível* o valor médio de uma série, num dado instante, descontados os efeitos dos outros componentes. Os gráficos da Fig. 4 mostram dois exemplos de séries que apresentam níveis aproximadamente constantes. O gráfico A mostra uma série de dados simulados; podemos notar que as observações oscilam em torno de um nível  $\mu=5$ . O gráfico B mostra uma série de medidas da variação mensal do nível do rio Amazonas, entre 1903 e 1992; a série oscila em torno de uma média nula.

Nas maioria séries reais, contudo, o nível não permanece constante, mas varia ao longo do tempo. Dizemos que uma série apresenta uma *tendência (trend)* quando seu nível aumenta ou diminui ao longo do tempo. Os gráficos da Fig. 5 mostram duas séries em que uma tendência crescente é bastante evidente.



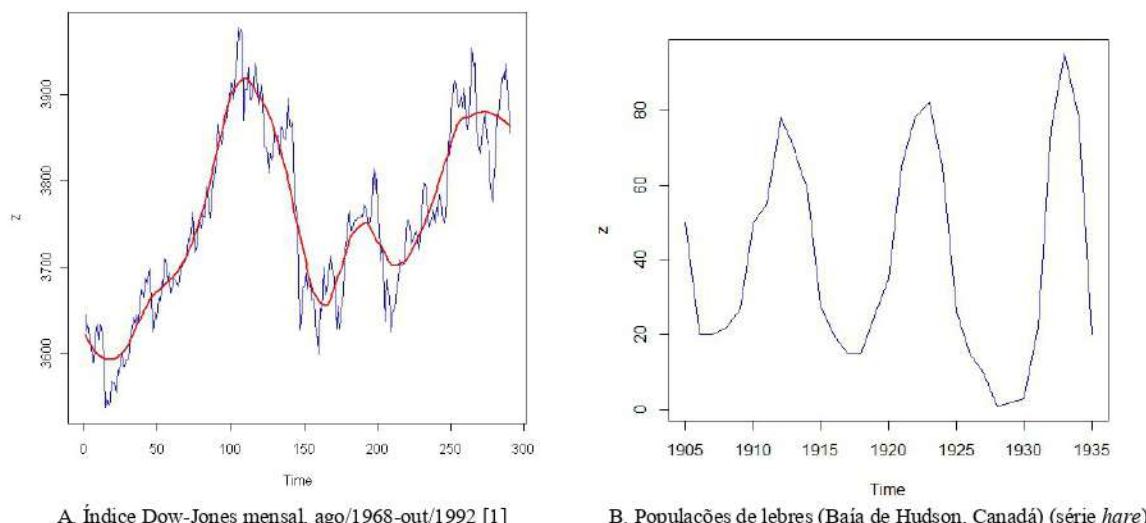
**Figura 4. Séries de nível constante**



**Figura 5. Séries com tendência crescente**

Há várias modelos matemáticos que podem ser usados para a tendência. Primeiro, há o modelo que considera o nível como função *linear* do tempo; poderia por exemplo ser aplicado em parte da série na Fig. 5B: de 1950 em diante, o nível sobe praticamente seguindo uma linha reta. Este modelo pressupõe que o nível é aumentado de um valor constante a cada ano. Na prática, é o modelo mais usado para previsões, pela sua simplicidade. Modelo um pouco mais complicado é o *quadrático*, que considera o nível como uma função de segundo grau do tempo; tanto este quanto o linear serão vistos no Cap. 4, aplicados a previsões que usam médias móveis. Outro modelo comum é o *exponencial*, às vezes usado para modelar o crescimento de populações, como o da Fig. 5A; este modelo pressupõe que o nível a cada ano é multiplicado por um fator constante. Na prática, nenhuma série pode crescer indefinidamente. Por isso, existem os modelos de tendência *amorte-cida*, que consideram que o aumento do nível a cada ano será uma fração constante do aumento do ano anterior; o aumento tenderá gradativamente para zero, e o nível irá se estabilizar a longo prazo. Às vezes também são usados modelos polinomiais de ordem superior a dois, curvas sigmoidais como a logística, etc.

Séries ligadas à Economia (como a da Fig. 5B) freqüentemente exibem um padrão cíclico – longos anos de crescimento, seguidos de quedas e recessões, outros anos de crescimento, etc. A Fig. 6A mostra a série de valores de fechamento mensal do índice Dow-Jones (bolsa de valores de Nova Iorque), exibindo ciclos de crescimento e decrescimento ao longo do tempo (a tendência, destacada em vermelho, foi estimada usando *loess*; seção 3.2.1.4). Estes padrões podem ser devidos aos chamados “ciclos de negócios” (*business cycles*), variações que ocorrem na economia de um país, de forma cíclica, porém irregular e imprevisível: anos de crescimento seguidos de anos de queda, etc. Existe porém muita discussão entre os economistas sobre ciclos e tendências; vários autores preferem não separar estes componentes, e falar em “tendência/ciclo” (*trend/cycle*). No restante deste livro, para simplificar, usaremos apenas a palavra ‘tendência’ para indicar a tendência / ciclo.



**Figura 6. Exemplos de séries mostrando ciclos**

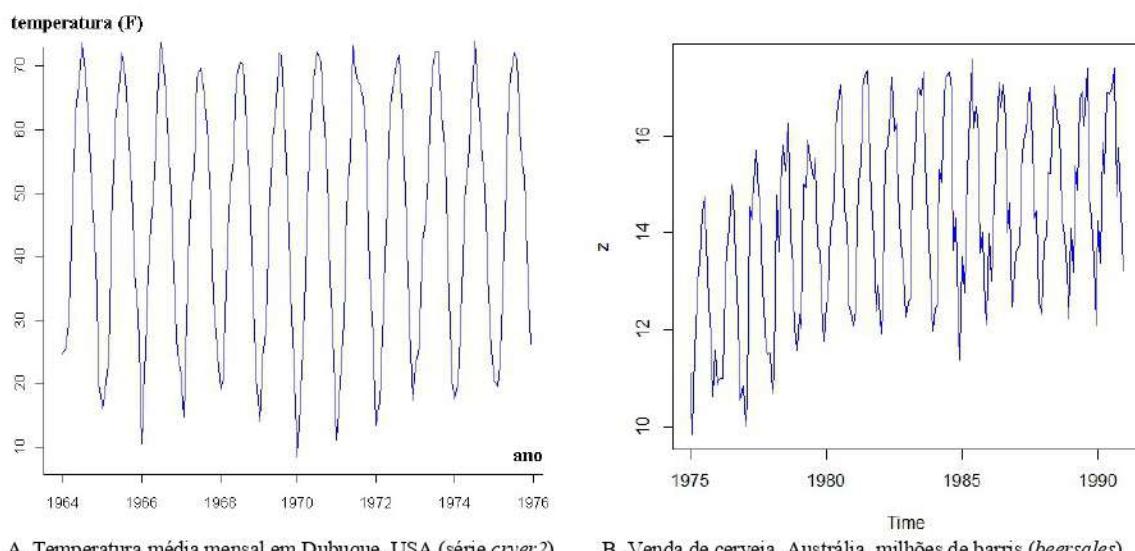
Estas variações cíclicas também ocorrem na natureza. A Fig. 6B mostra a população de uma espécie de lebres numa região do Canadá. A população cresce durante alguns anos, até ultrapassar a capacidade de seu meio ambiente, e depois passa a decair. Depois de

alguns anos, o ciclo se repete. Note que o ciclo tem duração irregular, e não deve ser confundido com a sazonalidade discutida na próxima seção. No gráfico é possível ver, por exemplo, que o último ciclo (1928 e 1935) teve duração mais curta do que o penúltimo (1917 a 1927).

### 2.2.3. Sazonalidade

Além das variações cíclicas irregulares mostradas na Fig. 6, que podem ocorrer ao longo de vários anos, uma série pode também exibir uma variação regular, que ocorre dentro de um intervalo de tempo fixo. Dizemos então que a série tem “sazonalidade” (*seasonality*), ou um comportamento “sazonal”. Por exemplo, suponha que uma loja de brinquedos queira fazer uma previsão de suas vendas em dezembro, para repor os estoques. Não faz sentido tomar como base as vendas do mês anterior (novembro) já que este é um mês fraco, em termos de vendas; melhor seria levar em conta as vendas de dezembro do ano anterior, devido à sazonalidade da série (isto é, ao fato de as vendas seguirem a cada ano um padrão regular de sobe-e-desce). No entanto, pode ser também necessário considerar a tendência das vendas durante o ano. Levando em conta o comportamento da economia durante o ano, esperamos que as vendas deste Natal sejam maiores, menores ou iguais as do Natal passado?

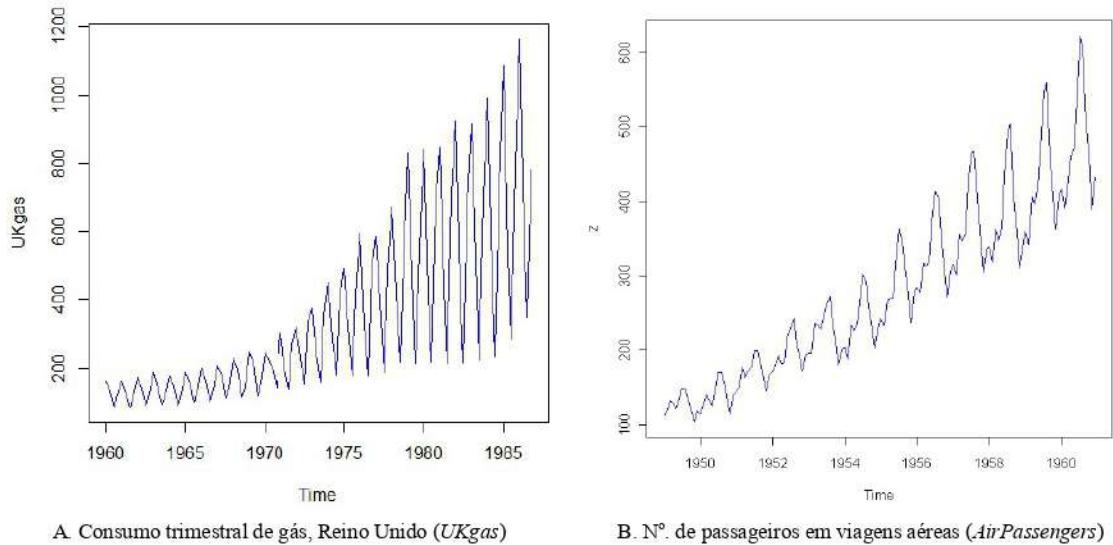
O gráfico da Fig. 7A mostra as temperaturas médias mensais em uma cidade. A série tem nível constante (isto é, não tem tendência), e uma sazonalidade anual óbvia; a cada ano, a temperatura alcança um máximo nos meses de junho ou julho, e um mínimo em dezembro ou janeiro, o que é o padrão das estações no hemisfério norte (note que o gráfico desta série se parece muito com o da série na Fig. 2B). A série de vendas de cerveja na Fig. 7B tem uma estrutura um pouco mais complicada. Além da sazonalidade (as vendas aumentam no meio do ano), houve uma tendência de crescimento linear até 1982, e depois disto o nível parece se estabilizar.



**Figura 7. Exemplos de séries sazonais**

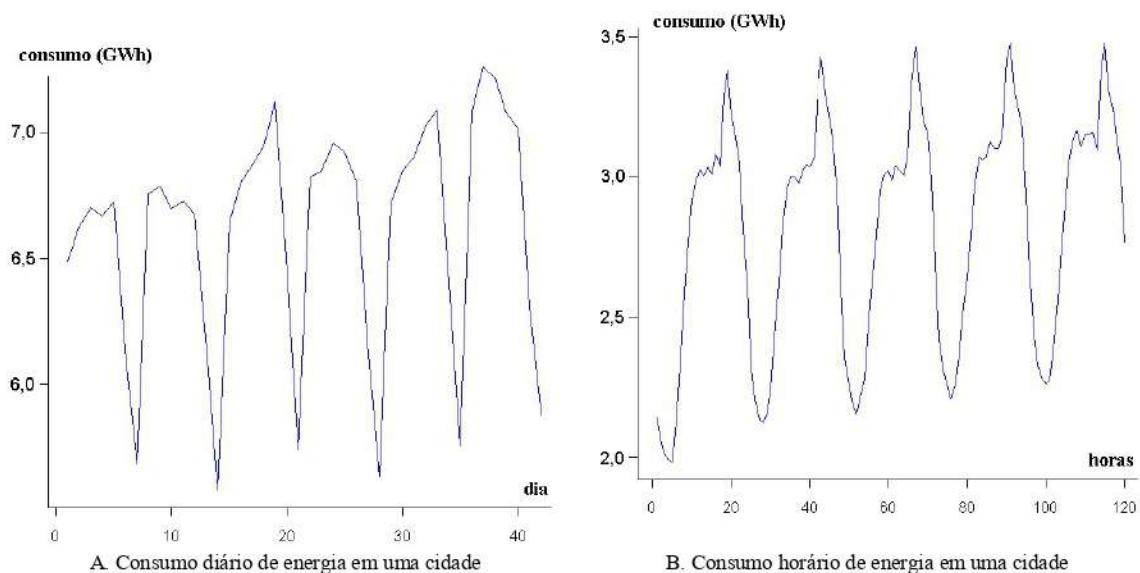
A julgar pelos gráficos, a amplitude da variação sazonal (isto é, a diferença entre o máximo e o mínimo a cada ano) é mais ou menos constante, nos dois exemplos anteriores.

Nas séries com forte tendência, porém, é mais comum encontrarmos ciclos sazonais de amplitude variável, como nas séries da Fig. 8. Nelas, a amplitude da oscilação cresce à medida que sobe o nível da série. Para trabalhar com tais séries pode ser necessário *transformar* os dados (veja Cap. 16).



**Figura 8. Exemplos de séries sazonais**

Na maior parte das vezes, a variação regular abrange o prazo de um ano; daí a palavra “sazonal”, derivada de “sazão”, que significa estação do ano (embora esta palavra seja pouco usada no Brasil; corresponde ao inglês *season* e ao francês *saison*). O número de intervalos que compõem um ano completo é chamado de *período* da sazonalidade, representado por  $s$ . Nos dois gráficos da Fig. 8, os valores foram medidos mensalmente, e o período é de  $s=12$ ; nos dois gráficos da Fig. 8, os dados são trimestrais e o período é  $s=4$ .



**Figura 9. Exemplos de séries com sazonalidade diferentes da anual**

Na análise de séries temporais, usaremos às vezes a palavra “sazonalidade” para indicar também comportamentos periódicos que ocorrem em intervalos diferentes de um ano. Por exemplo, o gráfico da Fig. 9A representa o consumo diário de energia elétrica em uma cidade, e mostra que há uma clara sazonalidade semanal ( $s=7$ ): o consumo é mais alto nos dias de semana (segunda a sexta-feira), e cai marcadamente nos fins-de-semana (sábado e domingo), uma vez que a maioria das indústrias, lojas e escritórios, estão fechados. A Fig. 9B, por outro lado, representa o consumo de energia elétrica a cada hora de dia, num intervalo de cinco dias (se segunda a sexta-feira); existe uma clara sazonalidade horária ( $s=24$ ), pois o consumo é mais alto durante o dia (com pico por volta das 18 h), e mais baixo durante a noite. Estes dois gráficos mostram que uma série pode vir a ter dois ou mais períodos sazonais sobrepostos (embora tais séries só sejam encontradas em algumas aplicações especiais).

#### 2.2.4. Resíduo ou erro aleatório

O último componente que podemos extraír de uma série temporal é o “resíduo” (*residual*), ou “ruído” (*noise*), ou “erro aleatório” (*error*). A denominação depende do tipo de modelo usado, e talvez da formação de quem estuda a série; engenheiros, por exemplo, tendem a chamar este componente de ruído, enquanto os estatísticos preferem chamá-lo de erro. Este componente é aleatório (devido ao acaso), de média nula, e imprevisível. É ele que faz com que técnicas estatísticas sejam necessárias para a análise e a previsão das séries; se não houvesse ruído, a previsão seria fácil – usariamos modelos determinísticos, como os usados na Física clássica. Numa série de nível constante, o nível de erro pode ser avaliado por meio de qualquer medida estatística de dispersão, como a *variância* ou o *coeficiente de variação*.

#### 2.2.5. Tipos especiais de séries temporais

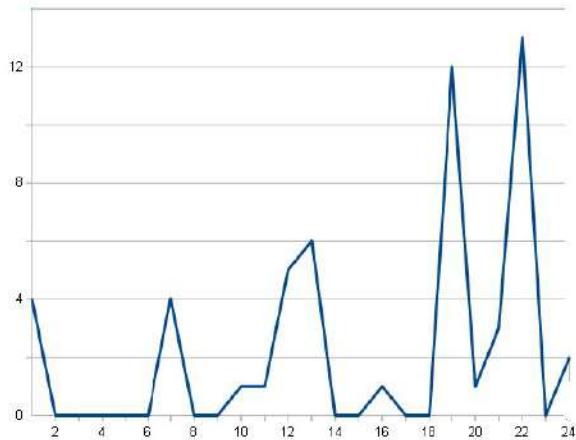
A maioria das séries encontradas podem ser analisadas em termos de um ou mais dos componentes vistos acima. Nas séries da Fig. 4, o nível é constante e a única variação existente é devida ao resíduo, bastante acentuado. Na Fig. 5, por outro lado, o que chama a atenção é a tendência, muito marcada, enquanto o resíduo é quase inexistente; na Fig. 5A, a série (pelo menos até cerca de 1930) parece ter sido gerada por uma função exponencial, que é um modelo determinístico. A série da Fig. 6A mostra uma tendência/ciclo bastante evidente, além de muito resíduo. A série da Fig. 7A tem apenas sazonalidade, nenhuma tendência, e pouco resíduo; a da Fig. 7B e as duas da Fig. 8 incluem, além da sazonalidade, uma forte tendência.

Existem ainda alguns tipos de séries que não podem ser analisadas por meio destes componentes. Veremos abaixo alguns destes tipos.

##### 2.2.5.1. Série de demanda intermitente (*intermittent series*)

O gráfico da Fig. 10 mostra a série de demandas mensais de uma peça de reposição, numa oficina de manutenção. Vemos que as demandas ocorrem de forma esporádica e parecem ser distribuídas de forma aleatória no tempo; não há nenhum padrão óbvio que mostre em que meses ocorrerão as demandas, nem qual será o total demandado. Este tipo de série é chamado por alguns autores de “série de demanda intermitente”. A previsão de uma série destas, se for possível, irá depender principalmente do conhecimento de como

funciona o mercado para este produto, quais são os compradores e quais são suas necessidades em função do estado geral da economia, etc.; não será possível prever esta série usando as técnicas usuais de séries temporais, como as discutidas neste livro.

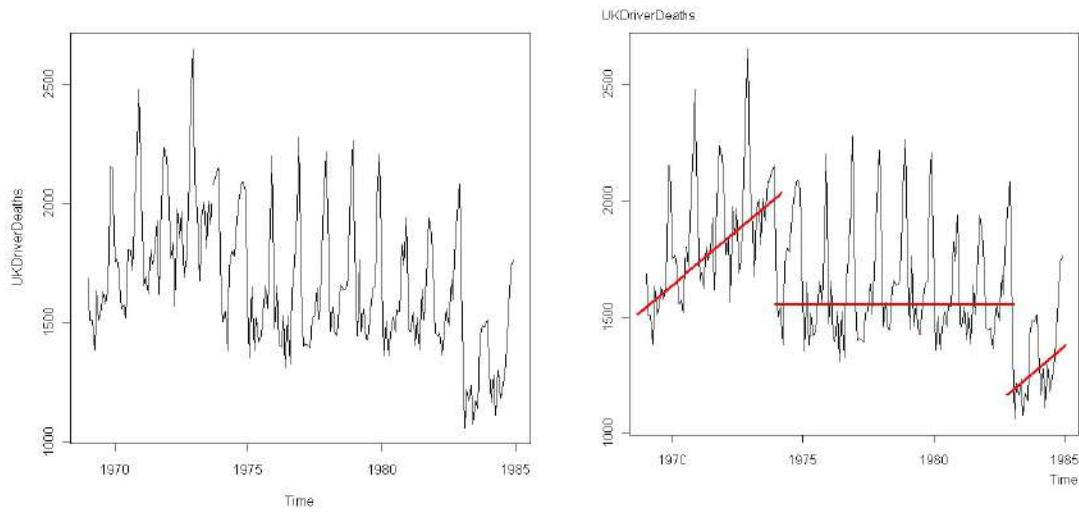


**Figura 10. Exemplo de série de demanda intermitente**

#### 2.2.5.2. Séries com quebras estruturais

No Cap. 1, mencionamos as séries que têm *quebras estruturais*, isto é, aquelas em que os padrões existentes se alteram repentinamente, por algum motivo; um exemplo mostrado foi a da Fig. 4 (Cap. 1), que ilustra a redução súbita da vazão do rio Nilo.

O gráfico da Fig. 11A mostra o número de mortos e feridos em acidentes de trânsito no Reino Unido, e é outro exemplo de série na qual a quebra estrutural é bastante evidente. A série tem, no início, tendência linear e sazonalidade (o valor cresce nos meses de início e fim de ano, que correspondem ao inverno). No ano de 1974, contudo, a série se altera bruscamente: a tendência desaparece, o nível cai de forma repentina, e permanece constante durante vários anos; a sazonalidade parece não ser muito afetada. Em 1983, ocorre outra quebra, e o nível cai novamente. A Fig. 11B mostra a variação do nível da série (estimada visualmente, de forma aproximada).



A. Mortos e feridos graves, entre motoristas, UK (SeatBelts)

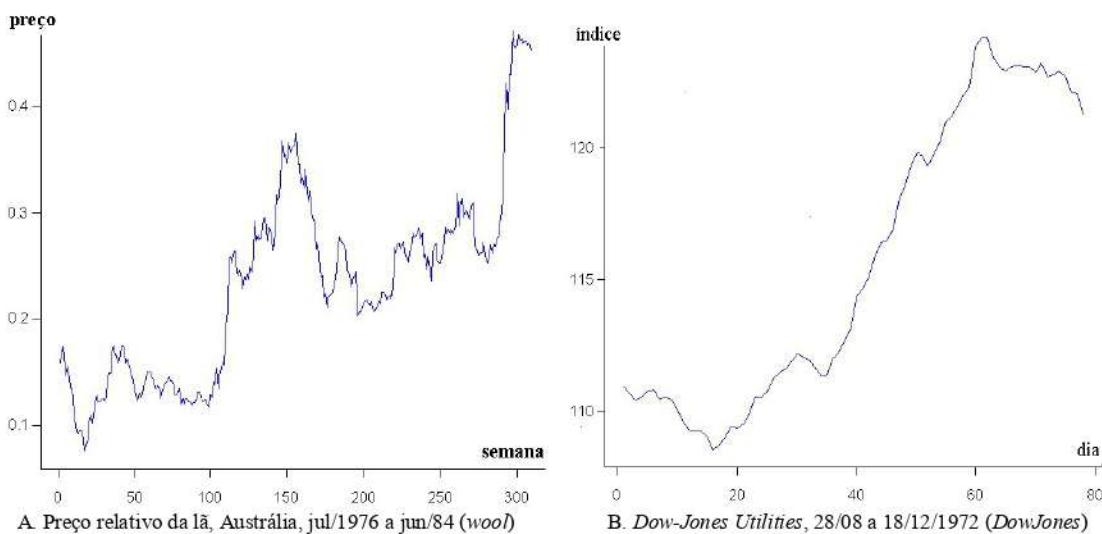
B. Comportamento do nível da série

**Figura 11. Exemplo de série com quebra estrutural**

Em geral, estas mudanças bruscas no comportamento de uma série são causadas por acontecimentos externos à série: crises econômicas, guerras, acidentes naturais, introdução de novas leis, etc. Estes acontecimentos (chamados de “intervenções”) são normalmente imprevisíveis. O que interessa ao analista é avaliar quantitativamente o impacto ocorrido na série, sabendo que uma intervenção ocorreu em uma data conhecida no passado. (No exemplo, a primeira intervenção foi o aumento repentino imposto pelos árabes ao preço do petróleo, em 1974, o que levou muitos países a reduzir o consumo, e a usar menos os automóveis; a segunda foi a introdução de lei obrigando os motoristas a usarem cintos de segurança, em fev/1983). Esta análise pode ser útil, por exemplo, para avaliar políticas públicas; na série acima, avaliar qual foi a redução percentual no número de mortes no trânsito, depois que os cintos de segurança se tornaram obrigatórios, e saber se esta redução será permanente ou apenas passageira. Exemplos destes modelos são os “modelos de intervenção” de Box e Jenkins (aplicações especiais dos modelos ARIMA, que veremos no Cap. 6); estes modelos contudo são complexos, e estão além do escopo deste livro.

#### 2.2.5.3. Séries de passeio aleatório

As séries da Fig. 12 não mostram nem tendência nem sazonalidade bem definidas; parece haver apenas erro aleatório, e um nível que parece variar de forma imprevisível. Este comportamento é observado com freqüência em séries financeiras. Um modelo possível para elas é o “passeio aleatório” (*random walk*; ver seção 10.4.5). O nome vem do traçado da série, parecido com o que seria produzido por um ponto que passeasse aleatoriamente no papel, para cima e para baixo. Este modelo implica, em essência, que estas séries são imprevisíveis, uma vez que não mostram nenhum padrão regular que possa ser extrapolado (há contudo muita discussão entre economistas e investidores sobre estes tipos de séries; para ver uma discussão, do ponto de vista de investidores, veja [2]).



**Figura 12. Exemplo de séries de séries financeiras**

#### Referências

- [1] *Time Series Data Library*. <http://www.robjhyndman.com/TSDL/>
- [2] Malkiel, Burton G. *A random walk down Wall Street*. NY: W. W. Norton & Co. 2007.