Capítulo 2. Conceitos e ferramentas básicas (cont.)

2.3. Medidas de erro de previsão

Suponha que tenhamos que prever os valores futuros de uma série, e que existam vários métodos de previsão disponíveis. Um dos critérios para escolher entre eles será a "qualidade": escolheremos o método que fizer previsões de maior qualidade. Na prática, é claro que existem também outros critérios que devem ser levados em conta, como o custo e a complexidade dos métodos; dependendo do problema, talvez não valha a pena usar um método muito caro e complicado, e pode ser melhor usar um método simples, mesmo que as previsões não sejam muito boas. Nesta seção, apresentaremos os conceitos e medidas mais usadas para avaliar a qualidade das previsões, além da terminologia e da notação que serão usadas ao longo do livro.

2.3.1. Terminologia e notação

Usaremos o símbolo Z_t para representar genericamente o valor de uma série temporal $Z_1, Z_2, ..., Z_T$. Usaremos a letra T para representar a último instante da série; em geral, irá corresponder ao instante atual. Chamaremos de *horizonte de previsão* (*lead time*) o número de instantes, ou de passos-à-frente, que separam o instante atual do instante para o qual queremos fazer a previsão, e o representaremos por k. A previsão de Z no instante $T+k$, feita a partir do instante T, será representada por

$$\hat{Z}_{T+k|T}$$

O chapéu ^ sobre o Z denota que se trata de uma *estimativa*, e não de um valor observado. Alguns livros preferem uma notação diferente:

$$\hat{Z}_T(k)$$

mas não a usaremos neste livro.

Chamaremos de *erro* ou *desvio* a diferença entre o valor Z_t que foi realmente observado num instante t, e valor $\hat{Z}_t$ que o método previu para esse instante:

$$e_t = Z_t - \hat{Z}_t \qquad (1)$$

Uma notação mais precisa deve indicar também o *horizonte* da previsão. Se estamos no instante t-1, e fazemos a previsão para o instante t, o erro desta previsão um-passo-à-frente será representado por

$$e_{t|t-1} = Z_t - \hat{Z}_{t|t-1} \qquad (2)$$

De forma geral, o erro da previsão k-passos-à-frente (isto é, da previsão de Z_t feita no instam-te t-k) é representado por:

$$e_{t|t-k} = Z_t - \hat{Z}_{t|t-k} \qquad (3)$$

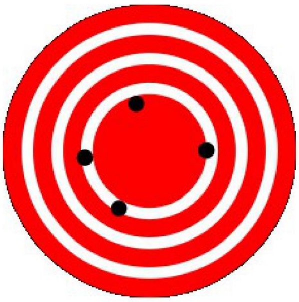
Neste livro, os erros de previsão considerados serão normalmente um-passo-à-frente, e usaremos com mais freqüência a notação simplificada em (**1**), em vez da forma em (**2**). A forma em (**3**) será usada quando for necessário distinguir erros de previsões com diferentes horizontes.

2.3.2. Acurácia e precisão

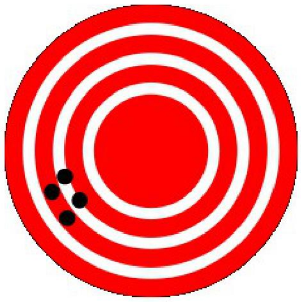
É claro que a qualidade de um método de previsão não pode ser avaliada a partir de uma previsão única; será preciso fazer uma série de previsões, e analisar os erros obtidos por meio de alguma forma de média. Em termos estatísticos, um bom método de previsão deve gerar erros que tenham duas características importantes:

1. Devem ter média que tende para zero. Se um método consegue estas previsões, dizemos que tem boa "acurácia", e que suas previsões são "não-tendenciosas".
2. Devem ter dispersão menor do que a dos erros dos outros métodos existentes. Se um método consegue estas previsões, dizemos que tem boa "precisão". [3]

Os conceitos de acurácia (*accuracy*) e precisão (*precision*) se originaram das áreas de ciências naturais e de engenharia, mas também são muito usados na Inferência Estatística. Uma analogia para explicá-los pode ser feita comparando um método de previsão com um atirador praticando tiro ao alvo. A Fig. **13** mostra os pontos atingidos no alvo, depois de algumas tentativas.



A. Resultados acurados, mas pouco precisos



B. Resultados precisos, mas pouco acurados

Figura 13. Acurácia e precisão (fonte: *Wikipedia*)

O atirador tenta acertar o centro do alvo; a distância entre o ponto que a bala atinge e o centro do alvo é o erro cometido. Se o atirador é *acurado*, os seus tiros se espalham de forma mais ou menos simétrica em torno deste centro; os erros em uma direção são compensados por erros na direção oposta. Se o atirador é *preciso*, as balas sempre atingem pontos próximos uns dos outros, não se espalhando muito pelo alvo. O ideal, é claro, é que os tiros sejam ao mesmo tempo precisos e acurados. Na APST, tentamos acertar a cada instante o valor real da variável, e o erro é a diferença entre este valor real e a estimativa que fizemos. Os erros sempre existirão; o ideal é que sejam pequenos (precisão), e em média, próximos de zero (acurácia).

Veremos abaixo alguns dos tipos de médias dos erros mais usados para avaliar a qualidade das previsões. Por enquanto, iremos nos ocupar apenas da média e da dispersão dos erros. Os conceitos a serem aplicados aqui são os mesmos que a Estatística Descritiva usualmente aplica a amostras aleatórias. Quando usamos métodos de base estatística, como os ARIMA, haverá outras exigências que os erros deverão atender, como as de não terem autocorrelação e terem distribuição normal; isto será visto no Cap. 14.

2.3.3. Medidas de erro de previsão baseadas em médias aritméticas

2.3.3.1. Erro médio (*ME*)

O Erro Médio (*mean error, ME*) é simplesmente a média aritmética dos erros cometidos numa seqüência de *N* previsões:

$$ME = \frac{1}{N}\sum_{t=1}^{N} e_t = \frac{1}{N}\sum_{t=1}^{N}(Z_t - \hat{Z}_t)$$

Em geral, esta medida não é muito útil. Se o método de previsão produz resultados não-tendenciosos, os erros positivos (valor observado maior que o previsto) irão compensar os erros negativos (valor observado menor que o previsto), e a média irá tender para zero. O ME pode servir para verificar se o método é *acurado*; não serve contudo para avaliar se ele é *preciso*, uma vez que não indica a dispersão dos erros.

2.3.3.2. Erro percentual médio (*MPE*)

O Erro Percentual Médio (*Mean Percent Error*, MPE) é uma medida baseada no valor relativo dos desvios, como o ME, mas apresentando os resultados em forma percentual:

$$MPE = 100 \times \frac{1}{N}\sum_{t=1}^{N}\left(\frac{e_t}{Z_t}\right) = 100 \times \frac{1}{N}\sum_{t=1}^{N}\left(\frac{Z_t - \hat{Z}_t}{Z_t}\right)$$

Como o ME, esta medida não é muito usada na prática, pois sempre a tende a zero se as previsões não forem tendenciosas.

2.3.3.3. Erro absoluto médio (*MAE*)

O Erro Absoluto Médio (*Mean Absolute Error, MAE*) é equivalente ao *desvio médio* (DM) usado na Estatística Descritiva: para evitar que os erros positivos compensem os negativos e a média se anule, calculamos a média dos *valores absolutos* destes erros.

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|e_t\right| = \frac{1}{N}\sum_{t=1}^{N}\left|Z_t - \hat{Z}_t\right|$$

As vantagens e desvantagens desta medida são as mesmas do DM: por um lado, o MAE é de interpretação muito simples; por outro, o uso do módulo dificulta o tratamento algébrico. Além disso, a função módulo dificulta os processos de otimização usados para ajustar os modelos, já que ela não tem derivada contínua.

2.3.3.4. Erro absoluto percentual médio (*MAPE*)

O Erro Absoluto Percentual Médio (*Mean Absolute Percent Error*, MAPE) é outra medida baseada nos desvios absolutos, como o MAE, mas apresenta os resultados em forma percentual:

$$MAPE = 100 \times \frac{1}{N}\sum_{t=1}^{N}\left|\frac{e_t}{Z_t}\right| = 100 \times \frac{1}{N}\sum_{t=1}^{N}\left|\frac{Z_t - \hat{Z}_t}{Z_t}\right|$$

Sua vantagem é a facilidade de interpretação: é evidente que um método que consegue erros de MAPE = 2% é mais preciso do que um que consegue erros MAPE = 5%. Esta medida é largamente usada em algumas áreas (por exemplo, na previsão de consumo de energia elétrica, cujos resultados são quase sempre relatados em termos do MAPE). Contudo, esta medida também tem alguns problemas. Primeiro, o uso do módulo dificulta o tratamento algébrico. Segundo, o MAPE não deve ser usados para séries que contenham observações iguais ou próximas de zero, porque estas observações irão aparecer no denominador, e podem causar problemas no processamento numérico.

Além disso, é preciso tomar alguns cuidados com a interpretação desta medida. O MAPE só faz sentido se a variável é medida numa *escala de razões* (i.e., o zero da escala tem significado absoluto). O MAPE de uma série de previsões de temperaturas em graus centígrados, por exemplo, será diferente do MAPE das mesmas previsões, se as temperaturas forem medidas em graus Fahrenheit. O MAPE também não faz muito sentido quando empregado em séries cuja média é baixa em relação à amplitude de oscilação da variável. Nestes casos, pode facilmente ocorrer que o MAPE tenha valor muito acima de 100%, mesmo que os erros sejam na realidade muito pequenos (por exemplo na série da Fig. **11** da seção 4.2.2).

2.3.3.5. Erro médio quadrático (*MSE*)

O Erro Médio Quadrático (*mean square erro, MSE*) é outra maneira de contornar o problema de os erros positivos compensarem os negativos, sem recorrer ao módulo: os erros são elevados ao quadrado, e calcula-se depois a média destes quadrados:

$$MSE = \frac{1}{N}\sum_{t=1}^{N} e_t^2 = \frac{1}{N}\sum_{t=1}^{N}(Z_t - \hat{Z}_t)^2$$

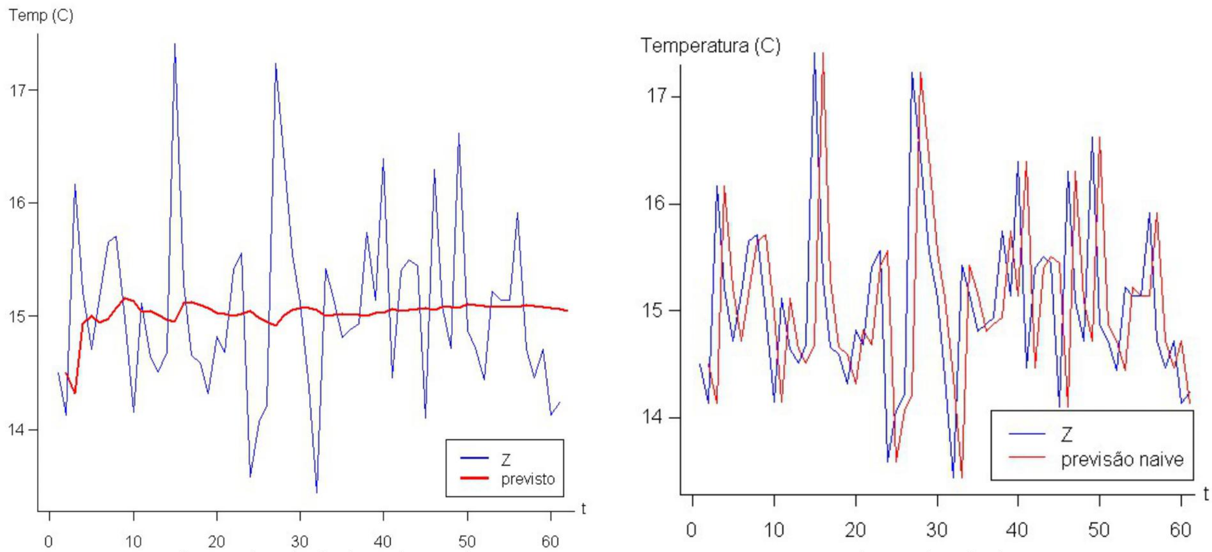
Esta medida é obviamente equivalente à *variância* usada na Estatística Descritiva, e é provavelmente a mais importante na APST. Vários dos procedimentos de estimação de parâmetros visam otimizar o MSE (por exemplo o método dos *mínimos quadrados ordinários*, seção 4.2.1).

Comparado com o MAPE, o MSE tem a desvantagem de ter como unidade o quadrado da unidade original da série e não permitir, por isso, uma interpretação intuitiva simples. Se fazemos por exemplo uma previsão de gastos de uma organização, em dólares, é fácil entender o que significa um erro de 3%; não é tão fácil entender o que significar um erro de um milhão de dólares ao quadrado. Além disso, por não se tratar de uma medida relativa, é difícil julgar a importância do erro; um erro de um milhão de dólares ao quadrado pode ser muito grande ou muito pequeno, dependendo da escala da organização.

Por outro lado, o MSE tem duas características vantajosas. Primeiro, ele pode ser usado em qualquer série, mesmo as que contenham zeros, ou que não sejam feitas numa escala de razões. Segundo, ele é muito sensível a valores discrepantes: os erros grandes encontrados na série de previsões terão bastante influência no valor do MSE, porque serão elevados ao quadrado. O MAPE é uma média simples dos erros absolutos cometidos; o MSE é uma média dos *quadrados* destes erros, e portanto penaliza mais o método que produzir erros grandes. Em problemas nos quais um erro grande tem conseqüências mais danosas do que vários erros pequenos, uma medida que penalize o método cujas previsões contenham ocasionalmente alguns erros grandes pode ser mais útil que uma medida que simplesmente dê o mesmo peso para todos os erros.

2.3.4. Método ingênuo (naïve) de previsão

As medidas de erro vistas acima servem para avaliar a qualidade das previsões feitas por um método. Contudo, não é possível decidir, com base apenas nelas, se uma série de previsões tem ou não qualidade aceitável. Suponha, por exemplo, que um método tenha conseguido previsões com um erro (MAPE) de 10 %. Estas previsões podem ser consideradas boas, ou não? Também não é possível decidir simplesmente comparando graficamente os valores previstos com os observados. A Fig. **14A**. mostra uma série de previsões de temperaturas, feitas por um método simples (*média simples*, que será visto no Cap. 4.2). As previsões não parecem grande coisa; parecem não acompanhar bem a variação da série.



A. Previsões pelo método de *média simples* B. Previsões pelo método *naïve*

Figura 14. Previsões de uma série de temperaturas médias anuais (série: *summer*)

Na verdade, não é possível avaliar a qualidade de um método de previsão se não houver um padrão com o qual ele possa ser comparado. Em geral, é usado como padrão de comparação o método *ingênuo* (*naïve*) ou métodos lineares simples.

O método ingênuo é a forma mais elementar possível de fazer uma previsão: supor simplesmente que o valor futuro será igual à observação mais recente, se a série for não-sazonal. Se a série tem valores diários, por exemplo, o método corresponde a dizer que o valor de amanhã será igual ao de hoje:

$$\hat{Z}_{t+1|t} = Z_t$$

Este método é normalmente chamado de método "ingênuo"; às vezes também se usa para designá-lo a palavra francesa *naïve*, que significa a mesma coisa. Num gráfico, a curva que representa a série de previsões ingênuas é idêntica à dos valores observados, apenas deslocada um instante para a frente; é o que pode ser visto na Fig. **14B**.

Se a série for sazonal, com período sazonal *s*, uma outra forma de previsão *naïve* é supor que o valor futuro será igual ao valor observado na mesma estação do ano passado:

$$\hat{Z}_{t+1|t} = Z_{t-s+1}$$

Por exemplo, suponha que uma série tenha observações mensais, com sazonalidade anual, e desejamos prever o valor do mês de março. Pelo primeiro método, a previsão *naïve* será o valor observado em fevereiro do mesmo ano; pelo segundo método, será o valor observado em março do ano passado.

Evidentemente, estas duas formas de fazer previsão são bastante primitivas, mas servirão para dar ao analista uma idéia de quão difícil é um problema, e que faixas de erro de previsão podem ser consideradas aceitáveis. Qualquer método de previsão proposto para uma série deve dar resultados melhores do que o método *naïve*; caso contrário, não faria sentido empregá-lo. Por exemplo, nas duas séries de previsão ilustradas na Fig. **14**, o previsor por média simples teve um erro MAPE de 4,1% contra os 5,1% do previsor *naïve*; os erros MSE dos dois métodos são, respectivamente, 0,68 e 1,17. Do ponto de vista do erro, portanto, o método da média simples pode ser considerado melhor que o *naïve*, neste problema. Outros exemplos do uso das medidas de erro serão mostrados depois que estudarmos os métodos de previsão, nos Capítulos 4 e 5.

Às vezes, contudo, a previsão ingênua é a única possível – o que corresponde no fundo a dizer que a série é imprevisível (isto acontece com séries que seguem o modelo conhecido como "passeio aleatório", cf. Cap. 10.4.5).

2.4. Tipos de métodos de previsão e análise de séries temporais

Existe uma grande variedade de métodos de análise e previsão; estudaremos nos próximos capítulos alguns dos mais freqüentemente usados. A não ser em alguns casos particulares (por exemplo, nos modelos econométricos, que são derivados de teorias pré-existentes sobre o comportamento dos sistemas econômicos), em geral não há regras que guiem a escolha do método a ser usado para resolver um determinado problema. Esta escolha pode ser guiada, em princípio, por critérios estatísticos: como em qualquer outra aplicação de modelos estatísticos, o melhor modelo é aquele que melhor se ajusta aos dados, ou que minimize uma função de custo escolhida. Na maioria das vezes, contudo, a escolha é influenciada também por fatores empíricos: o objetivo do trabalho, o tipo de informação disponível sobre o passado, o custo do erro, a acurácia necessária, o ambiente em que as previsões são feitas, a formação e o nível de conhecimento da pessoa que faz as previsões, etc. Discutiremos abaixo os tipos de métodos disponíveis, de acordo com o objetivo do trabalho.

2.4.1. Métodos para previsão

O objetivo do estudo pode ser simplesmente *prever* os valores futuros de uma série; para isto, são usados *modelos univaridos,* ou *modelos de séries temporais* (*time series models*), que fazem as previsões com base somente nas observações passadas da mesma série. Este tipo de modelo costuma ser denominado de modelos de *caixa-preta* (*black box*): o pesquisador conhece as variáveis de entrada (observações passadas da série) e as de saída (previsões), mas não sabe como funciona o mecanismo dentro da caixa.

Para a previsão, há dois grupos de métodos a considerar: os quantitativos e os qualitativos. Os métodos *quantitativos* (que este livro pretende discutir) são usados quando existe bastante informação disponível sobre o passado, e esta informação está em forma numérica. A série é analisada, e seus padrões são identificados (nível, tendência, sazonalidade, correlações); o que o método faz é extrapolar estes padrões para o futuro. (Para isto, estes métodos exigem a suposição de que pelo menos parte dos padrões observados no passado ainda serão mantidos no futuro). Os métodos *qualitativos* são menos acurados que os quantitativos, mas são o único recurso quando não há suficiente informação numérica. Estes métodos, que podem ser combinados com os quantitativos, são usados com mais frequência nas previsões para médio ou longo prazo, principalmente na construção de "cenários".

Os métodos quantitativos de previsão em geral se situam em algum lugar entre dois extremos. Num extremo, o método mais simples possível é o *naïve* (seção 2.3.4), que faz a previsão apenas repetindo a última observação disponível ("se no ano passado vendemos 10.000 peças, este ano também venderemos 10.000 peças"). Na prática, este método é ainda bastante usado em algumas organizações, com suas previsões sendo às vezes um pouco modificadas com base na experiência de quem faz a previsão ("no ano passado, vendemos 10.000 peças; este ano, porém, com o dólar subindo e o possível aumento da inflação, acho que as vendas irão diminuir uns 20%"). Existem muitos outros métodos heurísticos que foram desenvolvidos com base na experiência empírica dos usuários, e ainda são muito usados nas empresas – ou porque são simples, ou porque os técnicos envolvidos não conhecem as outras opções existentes. Um exemplos destes é o método de previsão do consumo de energia elétrica de uma região, um dia à frente, que era usado em empresas inglesas [[4]]: uma vez que o consumo depende da temperatura, o engenheiro examinava a série de temperaturas horárias previstas para o dia seguinte, e procurava no arquivo alguns dias do passado cujas temperaturas mais se assemelhassem a esta série; depois usava os consumos destes dias pas-sados como base para prever o consumo do dia seguinte.

Estes métodos informais contudo, são pouco acurados e, por não terem base estatística, não há como avaliar a confiabilidade de seus resultados. Além disso, eles não podem em geral ser *transportados* – quer dizer, foram criados para uso em uma situação, em uma organização, e não podem ser aproveitados para outro problema ou outro local. Devido à disseminação dos computadores e dos *softwares* de previsão, estes métodos tem sido aos poucos abandonados e substituidos por métodos mais formalizados. A previsão *naïve*, contudo, ainda tem um papel importante no estudo de séries temporais, pois serve de padrão para testar o desempenho de outros métodos: qualquer método mais complicado deve obter resultado melhores do que este, senão não vale a pena usá-lo.

No outro extremo, em termos de complexidade, estão os métodos de base estatística e os de inteligência computacional. Nos *métodos estatísticos*, a previsão é baseada na extrapolação, usando modelos probabilísticos conhecidos; exemplos destes são os modelos ARIMA (baseados na distribuição normal e na teoria de processos estocásticos), e os modelos estruturais. Ambos têm uma sólida base matemática, mas têm a desvantagem de não serem fáceis de explicar em termos intuitivos; seu mecanismo é compreensível apenas para quem tenha um conhecimento de Estatística razoavelmente sólido; além disso, existem que haja uma boa quantidade de dados (as séries usadas devem ser razoavelmente longas). O mesmo acontece com os modelos baseados em inteligência computacional, como as *redes neurais*; podem conseguir previsões muito acuradas, mas trabalhar com estes modelos exige bastante conhecimento especializado do usuário, e a quantidade de dados exigida é ainda maior do que a dos modelos ARIMA.

Entre estes dois extremos, há uma grande gama de métodos heurísticos (métodos de *médias móveis* e de *amortecimento exponencial*), que não exigem o ajuste de modelos probabilísticos e usam técnicas matemáticas básicas, como médias aritméticas simples ou ponderadas (mesmo que tenha sido demonstrado mais tarde que estes métodos não são mais que casos particulares de modelos ARIMA mais gerais). A simplicidade é um dos atrativos destes métodos, pois é relativamente fácil explicar aos usuários das previsões (gerentes e outros tomadores de decisão) o que foi feito. O "apelo intuitivo" de um método pode ser um elemento importante na escolha; é bem mais fácil para o tomador de decisões usar com proveito as previsões se tiver uma idéia, ainda que em termos gerais, da maneira como elas foram obtidas. Outros argumentos em favor destes métodos é a facilidade de uso, e a pouca exigência em termos de recursos computacionais; podem ser automatizados, são fáceis de modificar para resolver problemas particulares; não exigem *softwares* caros ou complicados, pois podem ser implementados em qualquer planilha eletrônica. Além disso, resultados de competições internacionais de previsão [5] mostraram que eles têm muito bom desempenho, na maioria das séries.

Na maioria das disciplinas sobre Séries Temporais oferecidos em cursos de graduação são estudados os métodos de *médias móveis* e de *amortecimento exponencial*, e os modelos ARIMA. Veremos os primeiros nos Caps. 4 e 5, e os segundos nos Caps. 6 a 15.

2.4.2. Métodos para análise

Outro objetivo de um estudo pode ser *analisar* uma série em seus componentes, ou *descrever* a relação da série com outras variáveis ou com eventos externos que podem afetá-la (por exemplo, descrever como as vendas de um produto são afetadas pela cotação do dólar, ou pela taxa de inflação; ou como o consumo de energia em uma cidade é afetado pela temperatura do ar).

Se o interesse é a *análise* de uma série temporal, pode ser usado o método tradicional de *decomposição clássica*, que analisa a série em seus componentes (nível, tendência, sazonalidade e resíduo), usando geralmente técnicas simples, baseadas em médias aritméticas. Este método será visto com no Cap. 3. Às vezes, é possível verificar como uma série de interesse é afetada, por outras séries ou por eventos externos, a partir de uma análise visual de gráficos, como foi feito no exemplo da Fig. **4** (série de vazões anuais do Nilo).

Freqüentemente, porém, os efeitos não são tão óbvios, é será preciso usar *métodos causais*, que analisam uma série em função de outras variáveis, e quantificam a interdependência entre elas. As ferramentas básicas destes métodos são modelos estatísticos de *regressão linear*, que descrevem uma série temporal Z dada como uma função linear de outras séries independentes, e avaliam quantitativamente o efeito das variáveis independentes sobre a variável de interesse. A previsão dos valores futuros da variável de interesse pode então ser feita com base nos valores observados ou previsões das variáveis independentes.

Estes métodos, no entanto, tendem a ser bastante complexos, pois os modelos de regressão usuais não podem ser aplicados diretamente sobre séries temporais, e têm que ser modificados (os modelos de regressão usuais exigem que as observações de uma amostra sejam *independentes* entre si, e a característica mais importante de uma série temporal é justamente a *dependência probabilística* entre suas observações). Estes modelos em geral são muito complexos, e não costumam ser abordados em cursos introdutórios de séries temporais; faremos uma introdução a alguns deles no Cap. 18.

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.
[3] Moore, David S. *Statistics – Concepts and Controversies*, 3rd ed. NY: W. H. Freeman & Co. 1991
[4] Hagan, MT; Behr, SM. The time series approach to short term load forecasting. IEEE Trans Power Syst, Vol. 2, No.3, Aug 1987
[5] https://en.wikipedia.org/wiki/Makridakis_Competitions#Second_competition,_published_in_1993