

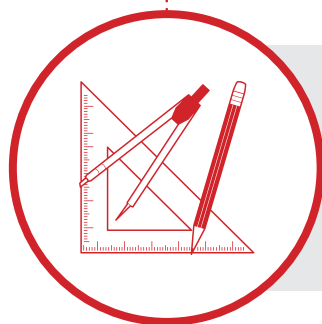


Matemática Multimídia

ANÁLISE DE DADOS
E PROBABILIDADE



GUIA DO PROFESSOR



Experimento


Séries temporais

Objetivos da unidade

1. Discutir e propagar, através de uma atividade, alguns conceitos de estatística;
2. Desenvolver a habilidade necessária para o tratamento de informações através de gráficos;
3. Introduzir aplicações práticas de estatística.



UNICAMP

LICENÇA Esta obra está licenciada sob uma licença Creative Commons 

Séries temporais

GUIA DO PROFESSOR

Sinopse

Com dados coletados diariamente durante um mês de alguma variável que evolui ao longo do tempo, os alunos construirão um gráfico e, analisando-o, tentarão prever o seu comportamento ao longo da próxima semana.

Conteúdos

Estatística, Interpretação de Gráficos e Dados.

Objetivos

1. Discutir e propagar, através de uma atividade, alguns conceitos de estatística;
2. Desenvolver a habilidade necessária para o tratamento de informações através de gráficos;
3. Introduzir aplicações práticas de estatística.

Duração

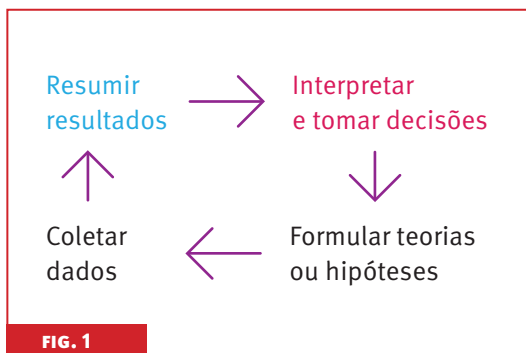
Uma aula simples.



Introdução

A pesquisa científica é um processo de aprendizagem. Neste contexto, os métodos estatísticos empregam ferramentas matemáticas que permitem otimizar este processo. Basicamente, podemos identificar os seguintes estágios:

1. Formulação de uma hipótese, que acarretará certas consequências;
2. Amostragem de dados;
3. Resumo, representação gráfica e comparação dos dados obtidos com o que era esperado, de acordo com a hipótese estabelecida;
4. Aceitação ou rejeição da hipótese. No caso de rejeição, formulação de uma nova hipótese; no caso de aceitação, conservação da hipótese até que novas amostras determinem sua rejeição.



Essas etapas formam um ciclo iterativo entre o avanço teórico (hipótese), que está relacionado ao levantamento de possíveis respostas para um problema específico, e os procedimentos de obtenção de dados.

Ao coletar dados, estamos interessados em obter informações que permitam manter a validade de uma hipótese ou que entreguem evidências suficientes para rejeitá-la, e então formular novas hipóteses, que serão testadas com uma nova coleta de dados, e assim por diante.

Motivação

Muitas das informações que recebemos em nosso dia a dia dizem respeito a dados numéricos que apresentam variabilidade no tempo. Em economia, por exemplo, temos o preço do Real em relação a alguma moeda estrangeira ou a taxa mensal de desemprego no país durante os últimos cinco anos. Em meio ambiente, temos as temperaturas média e máxima diárias em certa região durante o último mês ou os níveis de radiação observados. Em estudos sociais, podemos considerar a evolução da proporção de votantes em um certo candidato, e assim por diante.

Conjuntos de dados referentes a variáveis numéricas observadas durante um período de tempo são chamados *séries temporais*.

Este experimento visa realizar um estudo descritivo de um conjunto de dados obtido pelos alunos, com o fim de conhecer algumas das ferramentas utilizadas nesta área da análise estatística.

O experimento

Os alunos devem realizar a coleta dos dados durante um mês antes da aplicação em sala de aula. Os dados a serem usados serão coletados diariamente dos meios de comunicação disponíveis e podem ser da área de interesse do aluno.

Algumas sugestões de variáveis que podem ser observadas são: temperatura máxima, umidade relativa do ar, valor diário do preço das ações da Petrobrás na IBOVESPA ou cotação do dólar.

Uma tabela como a TABELA 1 do EXPERIMENTO deve ser preenchida pelos alunos com os dados obtidos. No exemplo, analisaremos a cotação diária do dólar durante os dias 3 a 31 de maio de 2010.

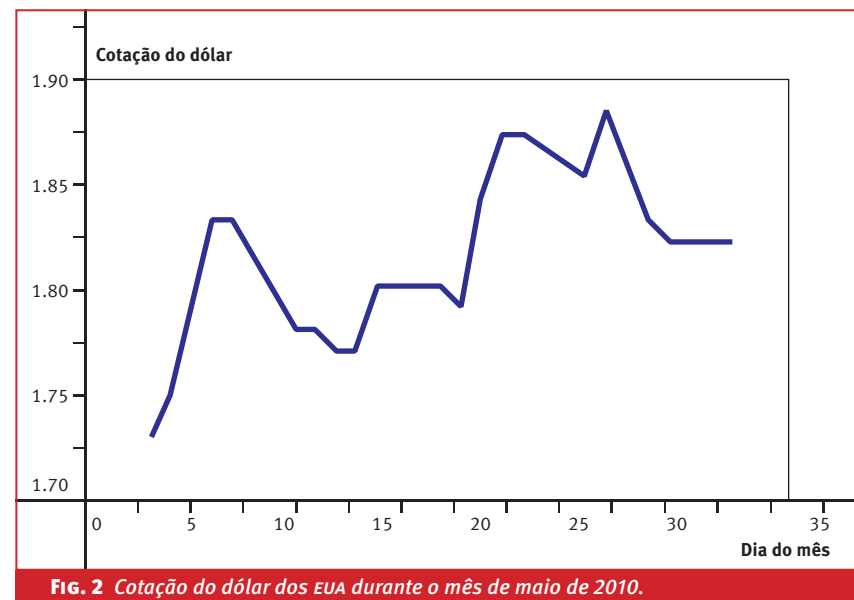
Dias do Mês (Maio/2010)	Cotação do Dólar (R\$)
3	1,73
4	1,75
5	1,79
6	1,83
7	1,83
10	1,78
11	1,78
12	1,77
13	1,77
14	1,80
17	1,80
18	1,79
19	1,84
20	1,87
21	1,87
24	1,85
25	1,88
26	1,85
27	1,83
28	1,82
31	1,82

TABELA 1 Cotação do dólar dos EUA durante o mês de maio de 2010.

Etapa 1 Os dados

Como primeira atividade em sala de aula, os alunos deverão fazer um gráfico representando os dados coletados.

Um gráfico de evolução temporal é construído adotando o tempo no eixo x e os valores da variável estudada no eixo y. Desta forma, para cada instante (no caso, para cada dia) há um único valor da variável associada. Usualmente, os pontos obtidos desta maneira são unidos por segmentos de reta. Para a tabela de dados apresentada no exemplo, o gráfico da série temporal é o da figura seguinte.



Observe que há uma tendência crescente no gráfico, com uma variação em torno de 10 centavos de real por cada dólar durante o período analisado, e uma variação máxima de 15 centavos.

À primeira vista, isso não parece muita diferença, mas se pensarmos em alguém querendo comprar uma passagem aérea de 1000 dólares, por exemplo, isso representará um aumento de 100 reais durante o período observado. Ou seja, se a pessoa não comprou a passagem no dia 3 e comprou no dia 25, por exemplo, ela perdeu 150 reais. Se estivermos falando de uma transação comercial ou de uma dívida em milhares ou milhões de dólares, esta diferença em reais é de fato significativa.

Planos de investimentos em dólares lidam exatamente com essas flutuações. Se uma pessoa tivesse aplicado 1000 dólares no dia 3 e os vendesse no dia 25, ela teria um ganho bruto de 150 reais.

Pela natureza dos dados, é razoável assumir que o preço do dólar em um certo dia dependa do preço do dólar nos dias anteriores. Uma pergunta possível aqui é: depende de quantos dias anteriores? Um dia anterior, dois dias anteriores, uma semana anterior? Podemos analisar esta pergunta à luz da informação entregue pelos dados, através das diferenças entre o preço do dólar de um dia para o outro.

O símbolo Δ_n representa a diferença entre o preço no dia n e o preço no dia $(n - 1)$, para n variando de 2 até o tamanho total da amostra. O seguinte gráfico mostra a série temporal das diferenças observadas.

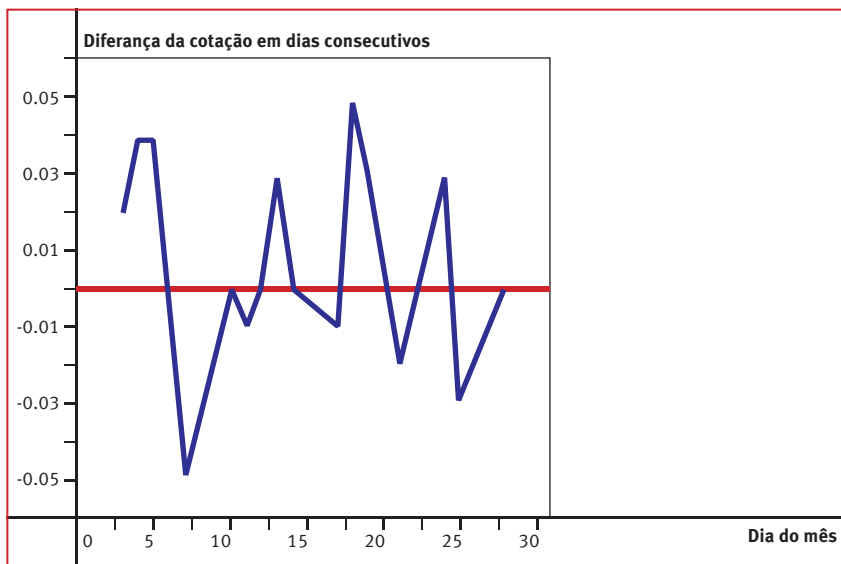


FIG. 3 Gráfico da série temporal das diferenças entre cotações do dólar em dias consecutivos.

Observemos que aparentemente as diferenças não apresentam nenhuma tendência no tempo e se distribuem de forma aleatória em torno do zero, ou seja, não há predominância de diferenças positivas ou negativas. Aparentemente, poderíamos considerar as diferenças como variáveis aleatórias independentes.

O seguinte gráfico de dispersão relaciona cada diferença com a diferença seguinte.

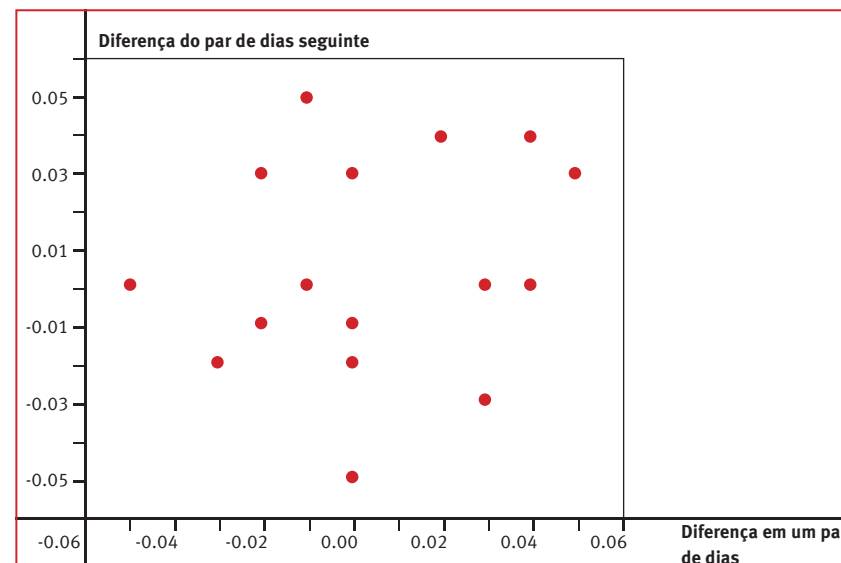


FIG. 4 Gráfico da diferença entre as diferenças.

Esses dois gráficos mostram uma grande evidência a favor da suposição de que as diferenças são independentes, já que a nuvem de pontos não apresenta qualquer tipo de relação entre as diferenças.

Esta suposição permite estabelecer um modelo matemático relativamente pouco complexo para a cotação do dólar, em que o valor de um dia pode ser previsto pelo valor do dia anterior e pela tendência (se houver) observada na série. Contudo, a construção desses modelos foge dos elementos conhecidos no Ensino Médio e, por isso, não serão apresentados aqui de forma mais rigorosa.

Etapa 2 Previsão da cotação do dólar

Nesta segunda etapa, os alunos deverão fazer sua própria previsão para os próximos valores da variável analisada. A previsão representa os valores esperados que eles poderiam observar nos dias seguintes ao da última medição feita na amostra.

Peça que eles façam uma previsão para os próximos 3 a 5 valores e que justifiquem como eles escolheram os valores previstos. No fechamento da atividade, os alunos poderão avaliar a qualidade de suas previsões de forma quantitativa e gráfica.

Consideremos as seguintes previsões e valores futuros observados no exemplo da cotação do dólar.

Dias do Mês (Junho/2010)	Previsão – Cotação do Dólar (R\$)	Cotação observada do dólar (R\$)
1	1,82	1,82
2	1,83	1,84
3 (Feriado)	–	–
4	1,83	1,84
7	1,84	1,86
8	1,85	1,87
9	1,87	1,84
10	1,84	1,82

TABELA 2 Valores da previsão e da cotação real do dólar dos EUA durante a semana seguinte da realização do EXPERIMENTO.

A segunda coluna contém os valores previstos e na terceira podemos encontrar os valores que foram de fato observados nos dias posteriores aos da amostra.

É possível representar esses valores no mesmo gráfico. A linha contínua indica os valores observados na semana seguinte e a linha tracejada aponta os valores de nossa previsão.

Uma medida usual para qualidade de uma previsão é o chamado erro quadrático médio, EQM da previsão. Este erro é de certa forma uma distância entre os valores observados e os valores previstos, e ele mede um erro médio da previsão. Intuitivamente, quanto menor o EQM melhor é a previsão.

Denotando por y_1, \dots, y_5 os cinco valores previstos pelo grupo e por d_1, \dots, d_5 os cinco valores observados na semana seguinte, o EQM é definido por:

$$EQM = \sum_{i=1}^5 \frac{(y_i - d_i)^2}{5}$$

No exemplo dado, fizemos uma previsão para sete dias. Assim, o EQM da previsão feita, igual à média entre os quadrados da diferença, é igual a 0,00031.

Para facilitar o cálculo do EQM, os alunos podem preencher uma tabela como a TABELA 5 do EXPERIMENTO, adicionando à tabela anterior mais duas colunas, uma com as diferenças entre o valor previsto e o valor observado, e outra com quadrado destas diferenças.

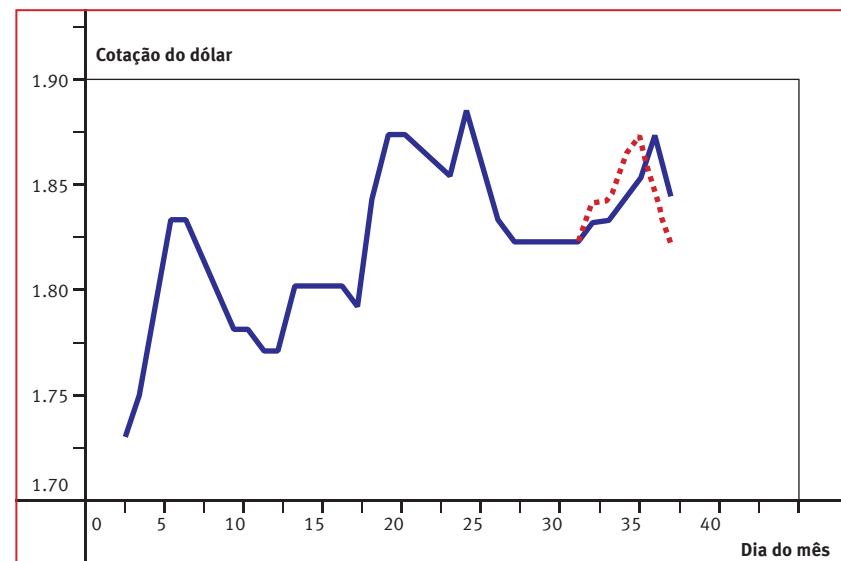


FIG. 4 Gráfico com os valores da previsão e da cotação real do dólar dos EUA.

Desta forma, é mais fácil corrigir possíveis erros de conta.

Uma boa previsão deveria ter resíduos positivos e negativos como se estivessem embaralhados, indicando que o modelo não tende nem a superestimar nem a subestimar os valores de interesse. Além, é claro, de ter resíduos pequenos em relação à variável estudada.

Fechamento

O foco deste experimento é o uso de tabelas e gráficos para tratar as informações coletadas. Discuta com os alunos a utilidade dessas ferramentas e as soluções obtidas por cada grupo.

É possível que surjam outras perguntas suscitadas pelo experimento, como as sugeridas no FECHAMENTO do EXPERIMENTO. Motive os alunos a tentar responder a essas questões e, se for necessário, planejar sua própria amostragem e analisar outros conjuntos de dados.

Variações

A variação natural deste EXPERIMENTO é a realização da pesquisa e a análise com outros conjuntos de dados referentes a variáveis numéricas observadas durante um período de tempo, ou seja, com outras séries temporais.

Os exemplos já citados na introdução do Experimento e na motivação deste GUIA são: o valor diário do preço das ações da Petrobrás na IBOVESPA, a variação de temperatura do Brasil nos últimos 30 dias, a temperatura máxima nos últimos meses, o crescimento da população de certo estado nos últimos anos, a proporção de votantes em um candidato durante as últimas pesquisas eleitorais, o nível do mar ou a taxa mensal de desemprego no país durante os últimos cinco anos.

Bibliografia

COSTA NETO, Pedro Luiz de Oliveira. **Estatística**. Editora Edgard Blücher, 2002.

FELLER, W. **Introdução à Teoria das Probabilidade e suas Aplicações**. Editora Edgard Blücher, 1976 (fora de catálogo).

MEYER, Paul. **Probabilidade: Aplicações à Estatística**. Livros Técnicos e Científicos Editora, 2003.

ROSS, S. **Probabilidade**. Bookman Co., 2010.

Bancos de dados

- Dados meteorológicos, hidrológicos e ambientais do Ministério da Ciência e Tecnologia: <http://satelite.cptec.inpe.br/PCD/>;
- Base de dados ambientais SONDA: <http://sonda.cptec.inpe.br/basedados/index.html>;
- Dados financeiros do Banco Central: <http://www4.bcb.gov.br/pec/taxas/port/ptaxnpesq.asp?id=txcotacao&id=txcotacao>.

Ficha técnica

AUTORA

Laura Leticia Ramos Rifo

REVISORES

Matemática

Samuel Rocha de Oliveira

Língua Portuguesa

Carolina Bonturi

Pedagogia

Ângela Soligo

PROJETO GRÁFICO

E ILUSTRAÇÕES TÉCNICAS

Preface Design



UNIVERSIDADE ESTADUAL DE CAMPINAS

Reitor

Fernando Ferreira da Costa

Vice-Reitor e Pró-Reitor de Pós-Graduação

Edgar Salvadori De Decca

MATEMÁTICA MULTIMÍDIA

Coordenador Geral

Samuel Rocha de Oliveira

Coordenador de Experimentos

Leonardo Barichello

INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA (IMECC – UNICAMP)

Diretor

Jayme Vaz Jr.

Vice-Diretor

Edmundo Capelas de Oliveira

LICENÇA Esta obra está licenciada sob uma licença Creative Commons 