



## IAG MASTER EM DESENVOLVIMENTO GERENCIAL 2006

Fundamentos de Economia de Energia  
Mônica Barros, D.Sc.  
Aula 3  
10/06/2006

monica@mbarros.com

1

## Uma visão gerencial da Estatística e Séries Temporais



### ■ Séries Temporais

- Interpretação da saída de ajustes de modelos
- Comparação de modelos
- Valores faltantes – como tratar?
  - Splines
- Data mining
- Transformação e tratamento dados

monica@mbarros.com

2

## Uma visão gerencial da Estatística e Séries Temporais



- Redes neurais
- Quando aplicar qual método para longo, médio, curto e curtíssimo prazo?
- Como medir/quantificar influência de variáveis – o conceito de elasticidade
- Cenários – “não são minha praia” mas vou tentar falar um pouquinho

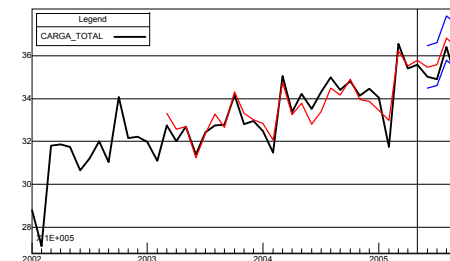
monica@mbarros.com

3

## Interpretação da saída de ajustes de modelos



### ■ Exemplo



Ajustamos um modelo de regressão dinâmica a esta série de carga total no período 03/2003 a 04/2005.

Quais os diagnósticos do modelo?

monica@mbarros.com

4

## Interpretação da saída de ajustes de modelos



### ■ Estrutura do Modelo

Forecast Model for CARGA\_TOTAL with log transform  
Regression(4 regressors, 0 lagged errors)

Term	Coefficient	Std. Error	t-Statistic	Significance
Log(CV_X3)	0.2059	0.0511	4.032	0.9995
Log(EMP_X11)	0.4197	0.1885	2.226	0.9639
Log(CARGA_TOTAL[-1])	0.1831	0.0750	2.441	0.9773
Log(CARGA_TOTAL[-12])	0.5401	0.1278	4.225	0.9997

## Interpretação da saída de ajustes de modelos



### ■ O que isso quer dizer?

### ■ O log da carga total depende das variáveis...

- Log(CV\_X3) – uma variável de consumo e vendas
- Log(EMP\_X11) – um indicador de emprego
- Log(Carga\_Total(-1)) – o log da carga um mês atrás
- Log(Carga\_Total(-12)) – o log da carga um ano atrás

## Interpretação da saída de ajustes de modelos



- Coeficiente – é o coeficiente estimado daquela variável no modelo
- Std. Error – é o erro padrão daquele coeficiente estimado
- T-statistic – é a estatística t. Se é maior ou igual a 2 em módulo indica que a variável é significativa e deve ser mantida no modelo.

## Interpretação da saída de ajustes de modelos



- Significance – é a significância da variável (ou do coeficiente). Quanto MAIOR a significância, MELHOR. O valor máximo da significância é 1.
- Alguns softwares exibem o “p-value” (p-valor), que é apenas 1- significância.
- Portanto, quanto MENOR o p-value, MELHOR.

## Interpretação da saída de ajustes de modelos



### Diagnósticos dentro da amostra

#### Within-Sample Statistics

Sample size	27	Number of parameters	4
Mean	15.03	Standard deviation	0.0397
R-square	0.8663	Adjusted R-squared	0.8488
Durbin-Watson	1.52	Ljung-Box(18)=16.14	P=0.4173
Forecast error	0.0154	BIC	61060
MAPE	0.0112	RMSE	47240
MAD	37470		

monica@mbarros.com

9

## Interpretação da saída de ajustes de modelos



Sample size	tamanho da amostra usada no ajuste do modelo
Mean	média da série
R-square	R-quadrado - quanto mais perto de 100% melhor
Durbin-Watson	Estatística que mede autocorrelação do erro de lag 1
Forecast error	Erro de previsão
MAPE	Erro médio PERCENTUAL de previsão
MAD	Erro médio ABSOLUTO de previsão

monica@mbarros.com

10

## Interpretação da saída de ajustes de modelos



Number of parameters	Número de coeficientes do modelo
Standard deviation	Desvio padrão da série
Adjusted R-squared	R-quadrado ajustado, penalizado pelo número de parâmetros
Ljung-Box(18)=16.14	Estatística de Ljung-Box - mede autocorrelações de resíduos nos primeiros k lags. Neste caso 18 lags. Idealmente, você NÃO QUER que esta estatística seja significativa.
BIC	Estatística BIC (Bayesian Information Criterion). Serve apenas para comparar modelos diferentes para a mesma série. Quanto menor, melhor.
RMSE	Raiz do erro quadrático médio.

monica@mbarros.com

11

## Interpretação da saída de ajustes de modelos



- Erros fora da amostra
- Neste exemplo “guardamos” os 5 meses finais da série para comparar os resultados das previsões com o efetivamente ocorrido.
- Os resultados são mostrados no Forecast Pro no “out of sample rolling evaluation”.

monica@mbarros.com

12

## Interpretação da saída de ajustes de modelos



Out-of-Sample Rolling Evaluation								
H	N	MAD	Cumulative Average	MAPE	Cumulative Average	GMRAE	Cumulative Average	Cumulative Average
1	5	56139	56139	0.016	0.016	0.475	0.475	0.475
2	4	69841	62229	0.02	0.018	3786	1.194	1.194
3	3	71882	64642	0.02	0.018	1113	1.174	1.174
4	2	88459	68044	0.025	0.019	0.663	1.082	1.082
5	1	24536	65144	0.007	0.019	0.353	1.004	1.004

H = horizontes de previsão (aqui de 1 a 5 meses à frente)

N = número de previsões feitas com aquele horizonte

MAD = erro absoluto médio

MAPE = erro percentual absoluto médio

monica@mbarros.com

13

## Interpretação da saída de ajustes de modelos



- Estes diagnósticos são importantes pois nos permitem verificar se as previsões se deterioram rapidamente com o aumento do horizonte de previsão.
- Por exemplo, neste caso vemos que nenhuma das estatísticas acumuladas está “explodindo”, o que indica que o modelo reage bem neste horizonte de 5 meses.

monica@mbarros.com

14

## Comparação de Modelos



- O que é um bom modelo?
- É algo que consegue estabelecer um balanço entre diversos requisitos, tais como:
  - Parcimônia (simplicidade)
  - Capacidade de ajuste dentro da amostra
  - Capacidade de gerar boas previsões
  - Interpretabilidade (coeficientes com sinais “certos”)
  - Inexistência de estrutura nos resíduos

monica@mbarros.com

15

## Comparação de Modelos



- As estatísticas de saída de um modelo nos permitem acessar a qualidade do modelo em relação a cada um destes aspectos.
- Por exemplo:
- Parcimônia – olhe para o  $R^2$  ajustado e o BIC.

monica@mbarros.com

16

## Comparação de Modelos



- Capacidade de Ajuste Dentro da Amostra
- Olhe para:
  - $R^2$  e  $R^2$  ajustado – quanto maior melhor
  - Erros – Forecast Error, MAD, MAPE e RMSE – quanto menor melhor
- Capacidade de Gerar Boas Previsões
- Olhe para:
  - “Out of sample rolling evaluation”

## Comparação de Modelos



- Inexistência de Estrutura nos Resíduos
- Olhe para:
  - Estatística Durbin-Watson – só autocorrelação de lag 1 dos erros. Se tudo estiver Ok, está perto de 2, mas é difícil dizer o quanto “perto” deve estar.
  - Estatística de Ljung-Box – calcula k primeiras autocorrelações dos resíduos, onde k depende do tamanho da série. Tem aproximadamente a distribuição Qui-quadrado. Idealmente é não significante.

## Comparação de Modelos



- Interpretabilidade
- Ajuste perfeito não adianta nada se os coeficientes não têm sinais coerentes. Por exemplo, preço explicando vendas com sinal positivo!
- Não faz sentido! Existe uma outra variável “por trás” que está sendo ignorada e que está mascarando o efeito do preço nas vendas – pode ser o próprio tempo.
- Não adianta ter um modelo com ajuste perfeito e coeficientes que não fazem sentido. Seja crítico!

## Comparação de Modelos



- Interpretabilidade
- No exemplo, as variáveis causais são:
- CV\_X3 = número de registros recebidos no SPC. Coeficiente positivo. Será que faz sentido?
- EMP\_X11=Média das horas efetivamente trabalhadas por semana, em todos os trabalhos, pelas pessoas de 10 anos ou mais de idade, ocupadas na semana de referência (Horas) . O coeficiente foi positivo. Neste caso parece fazer perfeito sentido para mim...

## Comparação de Modelos



Comparação de Modelos			
	Regressão Dinâmica	Holt-Winters	Box-Jenkins
N	27	41	15
R-quadrado	0.866	0.856	0.823
R-quadrado ajustado	0.849	0.848	0.818
Forecast Error	1.54%	2.20%	2.41%
MAPE	1.12%	1.57%	1.47%
MAD	37470	51420	48620
BIC	61060	79570	84420

Marcados em  laranja os melhores resultados dentre os 3 modelos.

monica@mbarros.com

21

## Valores Faltantes – como tratar?



- Depende...
- Existem algoritmos sofisticadíssimos e soluções simplérrimas.
- Algumas considerações importantes:
  - Quantidade de valores faltantes
  - “Localização” de valores faltantes – estão todos agrupados em blocos ou espalhados pela série?

monica@mbarros.com

22

## Valores Faltantes – como tratar?



- Soluções simples...
- Substituir por funções fáceis de calcular, por exemplo, médias e medianas.
- Soluções um pouco mais sofisticadas...
- Substituir valores faltantes por previsões obtidas através de métodos automáticos.

monica@mbarros.com

23

## Valores Faltantes – como tratar?



- Soluções um pouco mais sofisticadas...
- Estimar uma função de  $\sin(t)$  e  $\cos(t)$  e substituir os valores faltantes pelos valores da função.
- Usar uma interpolação linear ou um spline cúbico.
- Note que estas soluções (harmônicas e interpolações) são determinísticas, no sentido de não serem baseadas em modelos “estatísticos”.

monica@mbarros.com

24

## Valores Faltantes – como tratar?



- Spline cúbico
- É uma coleção de funções polinomiais de 3a. ordem que passam através de  $m$  “nós”.
- A segunda derivada de cada polinômio é geralmente igualada a zero nos extremos dos intervalos, o que produz um sistema de equações de contorno fácil de resolver para produzir os coeficientes dos polinômios.

## Valores Faltantes – como tratar?



- E como fazer para criar uma série “para trás” ?
- Por exemplo, a série de carga da Eletropaulo não existe antes de 1998. O que fazer? Pegar o “share” correspondente à sua área de concessão na série da Cesp antes de 1998 e construir uma série artificial que corresponderia à Eletropaulo antes de 1998.

## Data Mining



- Da Wikipedia...
- “Data Mining ou Mineração de Dados é um conjunto de técnicas que buscam a aquisição de novos conhecimentos através da análise de grandes bases de dados. Utilizam diversos algoritmos computacionais tais como Segmentação, Classificação e Previsão.”

## Data Mining



- Ainda da Wikipedia....
- “A premissa do Data Mining é uma argumentação ativa, isto é, em vez do usuário definir o problema, selecionar os dados e as ferramentas para analisar tais dados, as ferramentas do Data Mining pesquisam automaticamente os mesmos à procura de anomalias e possíveis relacionamentos, identificando assim problemas que não tinham sido identificados pelo usuário.

## Data Mining



- As ferramentas de Data Mining analisam os dados, descobrem problemas ou oportunidades escondidas nos relacionamentos dos dados, e então diagnosticam o comportamento dos negócios, requerendo a mínima intervenção do usuário ....”

## Data Mining



- Para nós em Séries Temporais ...
- É um bom “namoro” com os dados!
- Não é uma coleção de algoritmos automáticos.
- Olhe para os dados de diversas formas, tente extrair o máximo de informação possível contida neles.
- O gráfico da série é essencial!

## Transformação e Tratamento de Dados



- Os modelos estatísticos geralmente pressupõem que os dados são Normais (Gaussianos).
- Em linhas gerais, isso requer que eles sejam números reais, simétricos em relação à média.
- Como verificar? Faça um **histograma** dos dados, que é apenas um gráfico de barras com as frequências de ocorrência em diversos intervalos.

## Transformação e Tratamento de Dados



- O Excel, através da ferramenta de análise de dados, produz histogramas, como mostrado na próxima figura.
- Nota: o suplemento de Análise de Dados deverá ter sido previamente instalado.

# Transformação e Tratamento de Dados



The screenshot shows the Microsoft Excel interface with the 'Tools' menu open. The 'Data Analysis...' option is highlighted at the bottom of the menu. The spreadsheet data is visible in the background, showing a list of dates and electricity consumption values.

33

# Transformação e Tratamento de Dados



The screenshot shows the 'Data Analysis' dialog box in Microsoft Excel. The 'Histogram' option is selected in the 'Analysis Tools' list. The 'Input Range' is set to '\$B\$2:\$B\$328' and the 'Bin Range' is set to '\$E\$3:\$E\$9'. The 'Chart Output' checkbox is checked.

34

# Transformação e Tratamento de Dados



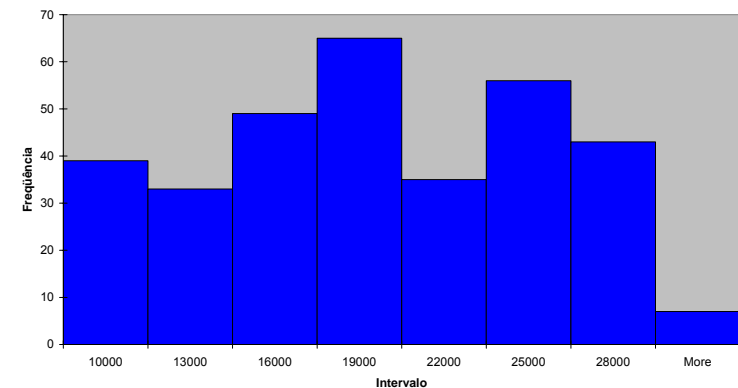
The screenshot shows the 'Histogram' dialog box in Microsoft Excel. The 'Bin Range' is set to '\$E\$3:\$E\$9', which corresponds to the 'Intervalos' column in the spreadsheet. The 'Chart Output' checkbox is checked.

35

# Transformação e Tratamento de Dados



Histograma - Consumo Total Brasil - 01/1979 a 03/2006



monica@mbarros.com

36

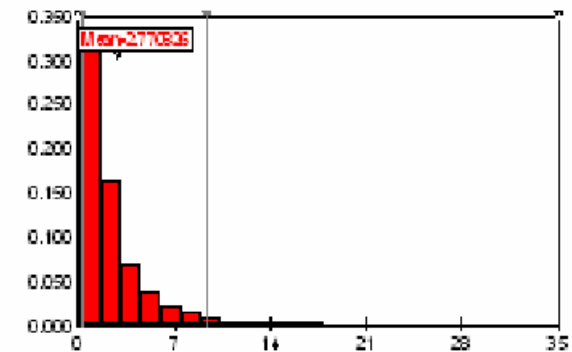


## Transformação e Tratamento de Dados

- Não tem uma “cara” muito Normal mas posso garantir que existem piores....
- O próximo gráfico contém o histograma de uma série que **CERTAMENTE** requer uma transformação de dados...



## Transformação e Tratamento de Dados

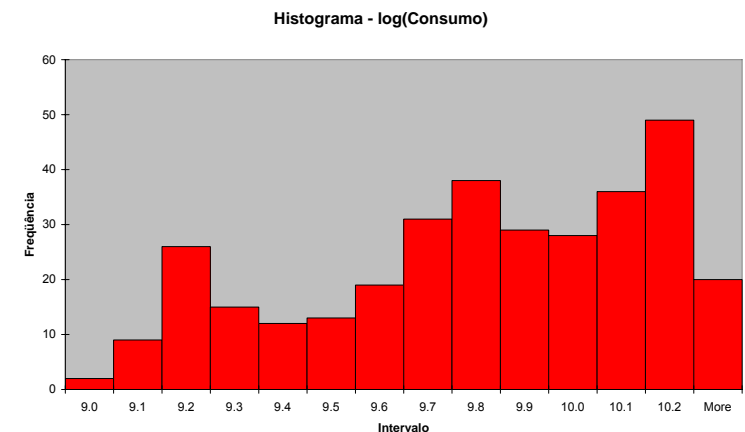


## Transformação e Tratamento de Dados

- Uma transformação usual para dados positivos (como o consumo de energia) é o logaritmo.
- A próxima figura apresenta o histograma do logaritmo do consumo. Nota-se que não houve uma melhora substancial na Normalidade dos dados.



## Transformação e Tratamento de Dados



## Redes Neurais



- As redes neurais artificiais (RN) são sistemas de processamento distribuído paralelo inspirados no sistema nervoso biológico.
- Como o cérebro humano, as RN são compostas de um conjunto de **unidades de processamento interconectadas**, chamadas de neurônios artificiais, ou apenas de **“neurônios”**.

monica@mbarros.com

41

## Redes Neurais



- RNs têm sido aplicadas em diversas áreas, como engenharia, ciências biológicas, economia, finanças, e computação.
- Algumas das mais importantes aplicações têm sido realizadas em **reconhecimento de padrões, aproximação de funções e previsão de séries temporais**.

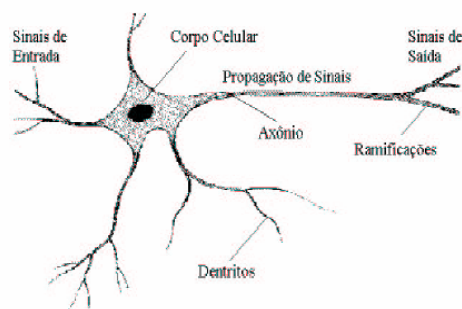
monica@mbarros.com

42

## Redes Neurais



### ■ Neurônio Biológico



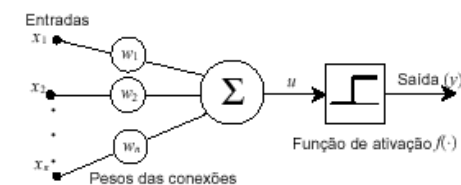
monica@mbarros.com

43

## Redes Neurais



### ■ Neurônio Artificial



O neurônio artificial desta figura é composto por  $n$  valores de entrada e um valor de saída (“output”)  $Y$ .

Os valores  $W_1, W_2, \dots, W_n$  são pesos aplicados aos “inputs” que resultam num potencial de ativação  $u$ .

monica@mbarros.com

44



# Redes Neurais

## ■ Neurônio Artificial

■ Este potencial de ativação  $u$  é então submetido a uma função de ativação  $f$  (geralmente não linear) e o **resultado final (a saída do neurônio) é  $Y = f(u)$** .

■ Em linhas muito simplificadas, o algoritmo anterior descreve o funcionamento de um neurônio em uma RN.



# Redes Neurais

■ No nosso caso específico ( $n$  inputs) então:

$$Y = f(u) = f\left(\sum_{i=1}^n W_i X_i\right)$$



# Redes Neurais

## ■ Funções de Ativação Mais Comuns

Sinal	$f(u) = \begin{cases} 1 & \text{se } u > 0, \\ 0 & \text{se } u \leq 0. \end{cases}$	
Rampa	$f(u) = \begin{cases} 0 & \text{se } u \leq \alpha, \\ \frac{u - \alpha}{\beta - \alpha} & \text{se } \alpha < u \leq \beta, \\ 1 & \text{se } u > \beta. \end{cases}$	



# Redes Neurais

## ■ Funções de Ativação Mais Comuns

Sigmoidal	$f(u) = \frac{1}{1 + \exp^{-\xi u}}$ <p>■ <math>\xi</math> é um parâmetro que determina o ponto de inflexão da função. A Figura mostra o caso particular <math>\xi = 2</math>.</p>	
Tangente hiperbólica	$f(u) = \tanh(\xi u) = \frac{\exp^{\xi u} - \exp^{-\xi u}}{\exp^{\xi u} + \exp^{-\xi u}}$ <p>Produzem saídas do neurônio positivas ou negativas. O gráfico mostra a função com parâmetro <math>\xi = 1</math>.</p>	



## Redes Neurais

### ■ O Neurônio Artificial Clássico

#### ■ McCulloch & Pitts (1943)

■ Estes neurônios são capazes de executar várias operações lógicas (OR, AND, etc.).

■ O neurônio clássico pode ser expresso por:

$$y = f(u) = f(x_1w_1 + x_2w_2 + \dots + x_nw_n) = f(w^T x)$$

■ Onde a função de ativação é binária e os  $x$ 's e  $W$ 's são como definidos anteriormente, e  $y$  é a saída do neurônio.

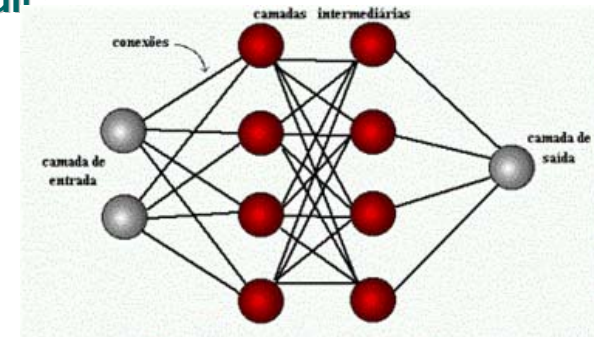
monica@mbarros.com

49



## Redes Neurais

■ Uma rede neural é uma interconexão de neurônios artificiais, como na figura a seguir:



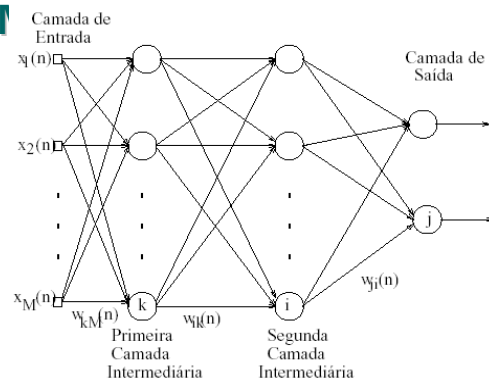
monica@mbarros.com

50



## Redes Neurais

### ■ Redes Perceptron Multicamadas (redes I)



monica@mbarros.com

51



## Redes Neurais

### ■ Redes Perceptron Multicamadas (redes MLP)

■ Consistem em:

- uma camada de entrada,
- uma ou várias camadas intermediárias e
- uma camada de saída.

monica@mbarros.com

52

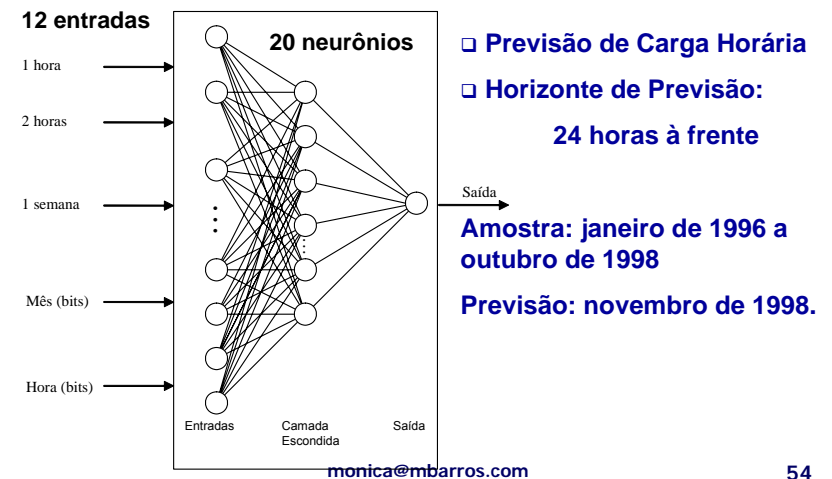


## Redes Neurais

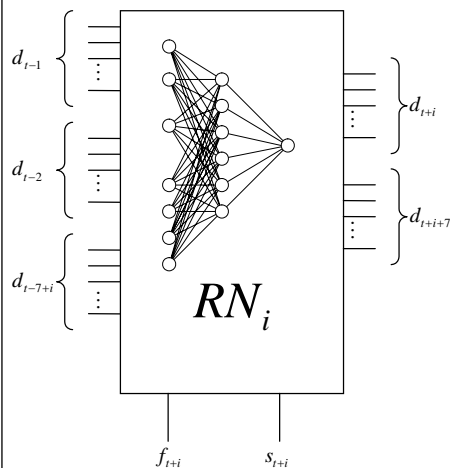
- As **camadas intermediárias** transmitem informações entre a camada de entrada e a camada de saída, e as funções de ativação dos neurônios desta camada são tipicamente funções de ativação não decrescentes e diferenciáveis, em geral **sigmóides**.



## Exemplo – Redes Neurais – Previsão de Carga Horária



## Exemplo - Rede Neural - Previsão 15 min



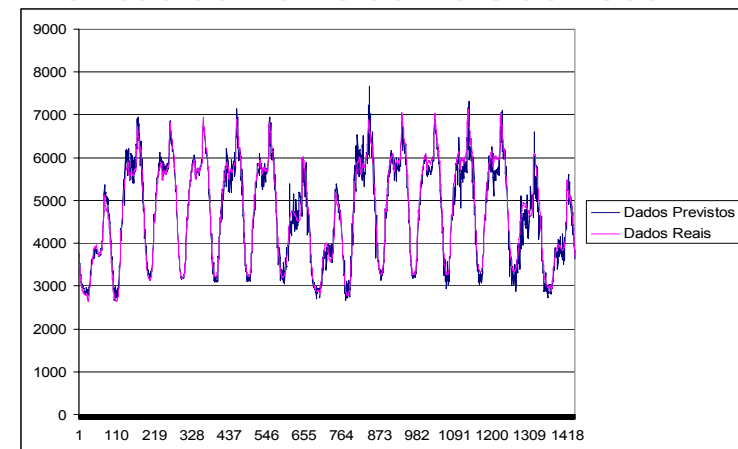
Rede que prevê o dia  $t+i$  (24 horas) e o dia  $t+i+7$  uma semana à frente usa como **informação**:

- dia  $t-1$  (ontem: 96 períodos de 15 minutos)
- dia  $t-2$  (ante ontem: 96 períodos de 15 minutos)
- dia  $t-7+i$  (7 dias antes do dia a prever)
- Horário de verão
- Dia da semana



## Exemplo RN – 15 minutos

- Previsão de 1 a 15 de Maio de 2005



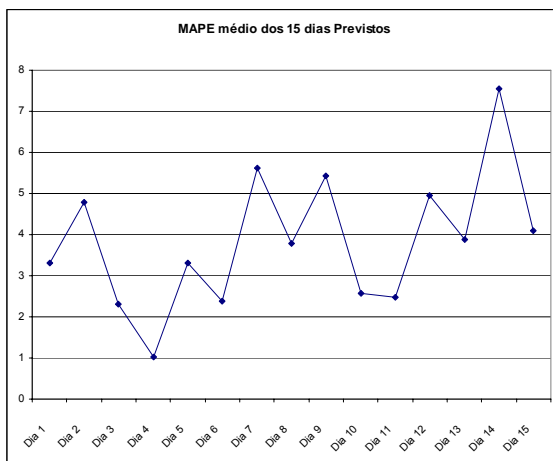
## Exemplo RN – 15 minutos



### ■ Previsão de 1 a 15 de Maio de 2005

MAPE Dias

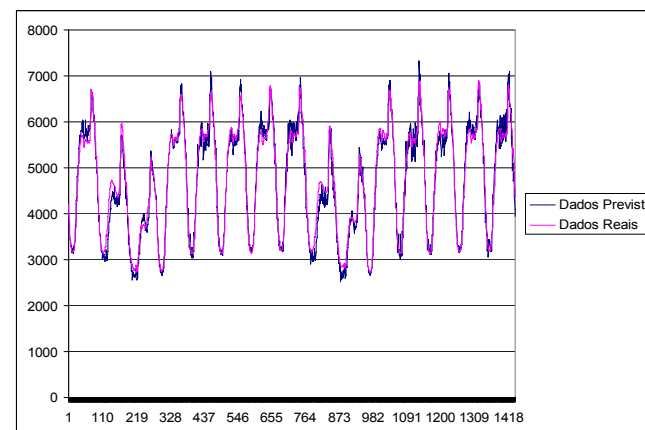
Dia	Erro
Dia 1	3.320026
Dia 2	4.780121
Dia 3	2.318703
Dia 4	1.02326
Dia 5	3.309194
Dia 6	2.383398
Dia 7	5.62829
Dia 8	3.793091
Dia 9	5.419788
Dia 10	2.568973
Dia 11	2.464699
Dia 12	4.944881
Dia 13	3.88723
Dia 14	7.554025
Dia 15	4.093368



## Exemplo RN – 15 minutos



### ■ Previsão de 15 a 29 de Julho de 2005



58

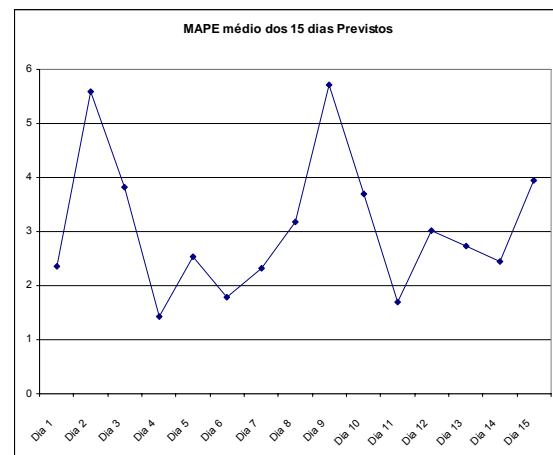
## Exemplo RN – 15 minutos



### ■ Previsão de 15 a 29 de Julho de 2005

MAPE Dias

Dia	Erro
Dia 1	2.366016
Dia 2	5.583273
Dia 3	3.813827
Dia 4	1.425647
Dia 5	2.541812
Dia 6	1.78931
Dia 7	2.328136
Dia 8	3.176319
Dia 9	5.710974
Dia 10	3.691544
Dia 11	1.704256
Dia 12	3.012898
Dia 13	2.728117
Dia 14	2.440695
Dia 15	3.946278



## Quando aplicar qual método?



- Curtíssimo prazo
- Periodicidade  $\leq$  1 hora
  - Redes Neurais
  - Modelos de Amortecimento Exponencial de Duplo Ciclo (Taylor)
- Curto Prazo
- Periodicidade Semanal
  - Modelos de Amortecimento Exponencial
  - Modelos Box-Jenkins

monica@mbarros.com

60

## Quando aplicar qual método?



- Médio Prazo
- Periodicidade Mensal
  - Modelos Causais
  - Modelos de Amortecimento Exponencial
  - Modelos Box-Jenkins
- Longo Prazo
- Periodicidade Anual
  - Modelos Causais
  - Cenários

## O conceito de elasticidade



### ■ Elasticidade

- Variação percentual na quantidade dividida pela variação percentual no preço (ou renda).
- No caso de consumo de um bem, a elasticidade-preço é negativa (quanto maior o preço, menor o consumo), e a elasticidade-renda é positiva (quanto maior a renda, maior o consumo).

## O conceito de elasticidade



### ■ Em termos algébricos:

$$\eta = \frac{\frac{\Delta Q}{Q}}{\frac{\Delta P}{P}}$$

← Variação percentual na quantidade

← Variação percentual no preço

## O conceito de elasticidade



### ■ No limite:

$$\eta = \frac{\frac{dQ}{Q}}{\frac{dP}{P}} \Leftrightarrow \frac{dQ}{dP} = \eta \frac{dP}{P}$$

Ao integrarmos esta última expressão encontramos:

$$\ln(Q) = \eta \cdot \ln(P) + const$$

A elasticidade é o coeficiente da variável explicativa "preço" numa regressão linear nos logs da quantidade e do preço.

## Cenários



- **Metodologia Utilizada no Plano Nacional de Energia 2030 (EPE) para criação de cenários macroeconômicos**
  1. **Análise do ambiente atual (nacional e mundial)**
  2. **Coleta de percepções e expectativas**
  3. **Identificação de condicionantes – tendências e incertezas**
  4. **Seleção de fatores críticos**
  5. **Formulação de hipóteses**
  6. **Geração de cenários exploratórios**
  7. **Quantificação**

## Cenários



1. **Análise do ambiente atual**
  - **Atualização dos trabalhos desenvolvidos em 2005 e aplicados no Plano Decenal 2006-2015.**
2. **Coleta de percepções e expectativas**
  - **Consulta a especialistas e extensão do horizonte dos cenários até 2030.**
3. **Identificação de condicionantes**
  - **Reconhecimento de tendências e incertezas.**

## Cenários



4. **Seleção dos Fatores Críticos**
  - **Definição das variáveis centrais.**
5. **Formulação de hipóteses**
  - **Geração de cenários exploratórios.**
  - **Definição de “cenas” (repartição temporal dos cenários).**
6. **Quantificação**
  - **Cenários Mundiais aferidos com referências disponíveis.**
  - **Cenários Nacionais aferidos através de modelos de consistência macroeconômica.**