

**Redes Neurais Aplicadas à Previsão de Demanda de
Numerário em Agências Bancárias**

Marco Antonio Freire Allemão

Universidade Federal do Rio de Janeiro
Curso de Mestrado

Orientador: Antonio Carlos Gay Thomé
Ph.D.

Rio de Janeiro
2004

Redes Neurais Aplicadas à Previsão de Demanda de Numerário em Agências Bancárias

Marco Antonio Freire Allemão

Dissertação submetida ao corpo docente do DCC/IM e NCE/UFRJ, Universidade Federal do Rio de Janeiro – UFRJ, como parte dos requisitos necessários à obtenção do grau de Mestre.

Aprovada por:

Prof. _____ - Orientador
Antonio Carlos Gay Thomé – Ph.D.

Prof. _____
Adriano Joaquim de Oliveira Cruz – Ph.D.

Prof. _____
Emmanuel Pisceces Lopes Passos – D.Sc.

Prof. _____
Nelson Francisco Fávilla Ebecken – D.Sc.

Rio de Janeiro
2004

Allemão, Marco Antonio Freire

Redes neurais aplicadas à previsão de demanda de
numerário em agências bancárias / Marco Antonio Freire
Allemão. - Rio de Janeiro, 2004.

xxi, 247 f.: il.

Dissertação (Mestrado em Ciência da Computação) -
Universidade Federal do Rio de Janeiro - UFRJ, Instituto
de Matemática - IM / Núcleo de Computação Eletrônica -
NCE, 2004.

Orientador: Antônio Carlos Gay Thomé

1. Previsão de Demanda de Numerário – Teses. 2. Séries
Temporais - Teses. 3. Redes Neurais – Teses. I. Antônio
Carlos Gay Thomé (Orient.) II. Universidade Federal do
Rio de Janeiro. Instituto de Matemática / Núcleo de
Computação Eletrônica. III. Título.

Às minhas queridas filhas Carolina e Raquel.

AGRADECIMENTOS

Muitos foram os que contribuíram de forma decisiva para a conclusão deste trabalho. Agradeço, sinceramente, a todos.

Entretanto, não poderia deixar de mencionar o meu amigo e chefe Marcos dos Anjos Teixeira, pelo seu imenso apoio nas incontáveis vezes que necessitei ausentar-me das minhas tarefas profissionais para dar prosseguimento ao curso.

À Cristiane Freire pelo seu companheirismo, incentivo e apoio. Por sua compreensão com meus estudos e auxílio nas revisões.

Ao Wilson Pumar, por seu compromisso com a formação acadêmica. Sem sua teimosia para permitir que eu pudesse realizar o curso, nada disso teria acontecido.

Aos grandes amigos José Olavo Marinho, Carlos Eduardo Andrade, Ricardo Moraes de Campos, Marcelo Vasconcellos, Antônio Araújo e a turma da Tesouraria pelo incentivo, ajuda, explicações e fornecimento do material de pesquisa.

Ao Prof. Thomé, pelo seu profissionalismo, dedicação e amizade. À sua postura de sempre incentivar e estimular novas pesquisas e descobertas.

À Lúcia Brantes pelo apoio e estímulo. Sem sua ajuda não teria conseguido percorrer este caminho.

À minha irmã Ana Valéria e ao professor Luiz Monteiro, pelo incentivo e vibração quando das minhas conquistas.

À prof. Lígia Barros pelos ensinamentos transmitidos, pela coragem em mudar o mundo e por ser, realmente, uma educadora.

Ao professor Sérgio Côrtes pelo auxílio despretenso, pelas valiosas observações e dicas.

Aos colegas de mestrado, Eugênio da Silva, Cracky (Roberto Rodrigues), Bruno Guingo e Alan Rezende e demais amigos do Laboratório de Inteligência Computacional - LabIC, pelo apoio e amizade.

RESUMO

ALLEMÃO, Marco Antonio Freire. **Redes Neurais Aplicadas à Previsão de Demanda de Numerário em Agências Bancárias..** Orientador: Antonio Carlos Gay Thomé. Rio de Janeiro: UFRJ/DCC-NCE, 2004. Dissertação (Mestrado em Ciência da Computação).

A cada dia inúmeras viagens são realizadas para transporte de valores entre as dependências de um banco. O objetivo é suprir/recolher dinheiro (numerário no meio técnico) referente ao movimento das agências com o público. Dessa forma, as empresas bancárias do mundo inteiro procuram balancear corretamente a quantidade de dinheiro a permanecer em suas filiais, de forma a não permitir ociosidade de recursos e, por outro lado, não deixar a agência desprovida de dinheiro. As instituições buscam modelos capazes de prever a necessidade futura, a fim de otimizar os custos com a operacionalização do processo. Esta dissertação procura desenvolver um modelo neural que prediga o nível ideal de recursos que uma dependência deve ter à sua disposição para realizar seus trabalhos do dia-a-dia. Realizou-se uma análise dos dados disponíveis e pesquisou-se a melhor arquitetura e topologia de rede para o problema. O modelo foi validado com algumas técnicas estatísticas usadas para predição de séries temporais, alcançando o melhor resultado.

ABSTRACT

ALLEMÃO, Marco Antonio Freire. **Redes Neurais Aplicadas à Previsão de Demanda de Numerário em Agências Bancárias..** Orientador: Antonio Carlos Gay Thomé. Rio de Janeiro: UFRJ/DCC-NCE, 2004. Dissertação (Mestrado em Ciência da Computação).

Every single day, money has to be taken back and forth around bank agency premises thousands of times a day. The purpose of that is to supply or collect cash derived from customers needs at bank agencies. Therefore, banking institutions all over the world try to correctly calculate the amount of money to be kept in each of its branches so as not to waste resources nor leave agencies running short of cash. These institutions search for models which would predict future needs in order to optimize costs when carrying out this process. This dissertation tries to develop a model which will foretell the ideal level of resources an agency must have at hand to perform its daily routine tasks. An analysis of the data regarding three agencies has been made and the best method to be applied to solve the problems has been researched. Both classic (statistical) and non-classic (neural networks and fuzzy logic) models have been dealt with. Results have then been analyzed and compared.

LISTA DAS FIGURAS

Figura 1: Figura representativa das fases do projeto: fase 1 => tratamento dos dados; fase 2 => Modelagem e treinamento da rede; fase3 => Avaliação do resultado e fase 4 => Comparação com modelos estatísticos.....	69
Figura 2: Modelo genérico do neurônio artificial	102
Figura 3: Modelo do Neurônio.....	102
Figura 4: Representação esquemática do Perceptron.....	104
Figura 5: Modelo do Adaline	104
Figura 6: Modelo de neurônio utilizando o back-propagation.....	105
Figura 7: Fluxo de processamento do back-propagation	106
Figura 8: Modelo de uma RNA do tipo Elman.....	107
Figura 9: Representação do resultado de diversos treinamentos: à esquerda tem-se o underfitting, ao centro um treinamento correto e à direita o overfitting.	112
Figura 10: Representação de um neurônio artificial.	116
Figura 11: Representação da superfície de erro com mínimos locais e globais.....	119
Figura 12: Representação da PCA e CPD. Na figura à esquerda tanto a PCA como a CPD conseguem discriminar as classes. Na figura (b) somente a CPD consegue definir a separação das classes.....	124
Figura 13: Neurônio artificial com parâmetros do algoritmo de back-propagation (THOMÉ, 2000, p. 16).	127
Figura 14: Gráficos mostrando como aumentar a precisão da aproximação nebulosa.....	136
Figura 15: Agência 02 – Separação dos dados em 4 intervalos.....	171
Figura 16: Estratégia de treinamento com o uso de médias móveis.	176
Figura 17: Representação da normal com os limites críticos para o nível de significância 0,01.....	191
Figura 18: Representação da normal com os limites críticos para o nível de significância 0,01.....	193

Figura 19: Representação da normal com os limites críticos para o nível de significância 0,01 e 0,05.....	195
---	-----

LISTA DOS GRÁFICOS

Gráfico 1: Agência 02 – representação de toda a série.....	42
Gráfico 2: Agência 06 – representação de toda a série.....	43
Gráfico 3: Agência 13 – representação de toda a série.....	43
Gráfico 4: Agência 02 - Valores mensais, médias e desvio padrão. O losango representa a média aritmética dos pontos no eixo das ordenadas (y) e as linhas horizontais correspondem aos valores referentes a 1 desvio e 2 desvios padrões da série.....	44
Gráfico 5: Agência 06 - Valores mensais, médias e desvio padrão. O losango representa a média aritmética dos pontos no eixo das ordenadas (y) e as linhas horizontais correspondem aos valores referentes a 1 desvio e 2 desvios padrões da série.....	45
Gráfico 6: Agência 13 - Valores mensais, médias e desvio padrão. O losango representa a média aritmética dos pontos no eixo das ordenadas (y) e as linhas horizontais correspondem aos valores referentes a 1 desvio e 2 desvios padrões da série.....	45
Gráfico 7: Agência 02 – Média +/- desvio padrão.....	46
Gráfico 8: Agência 06 – Média +/- desvio padrão.....	47
Gráfico 9: Agência 13 – Média +/- desvio padrão.....	47
Gráfico 10: Agência 02. Dias da semana e médias.....	49
Gráfico 11: Agência 06. Dias da semana e médias.....	50
Gráfico 12: Agência 13. Dias da semana e médias.....	51
Gráfico 13: Agência 02. Dias úteis do mês e médias.	52
Gráfico 14: Agência 06. Dias úteis do mês e médias.	54
Gráfico 15: Agência 13. Dias úteis do mês e médias.	55
Gráfico 16: Ag. 02 – Gráfico com os pontos no teto de 2 desvios padrões e com marcação dos outliers.	82
Gráfico 17: Ag. 06 – Gráfico com os pontos no teto de 2 desvios padrões e com marcação dos outliers.	83

Gráfico 18: Ag. 13 – Gráfico com os pontos no teto de 2 desvios padrões e com marcação dos outliers.	83
Gráfico 19: Agência 02 - $\sqrt{\text{abs}(\text{reg})}$	85
Gráfico 20: Agência 06 - $\sqrt{\text{abs}(\text{reg})}$	86
Gráfico 21: Agência 13 - $\sqrt{\text{abs}(\text{reg})}$	86
Gráfico 22: Modelo sazonal multiplicativo.	94
Gráfico 23: Modelo sazonal aditivo.....	95
Gráfico 24: Função mapeada pela rede neural. Dentro da elipse têm-se os valores previstos.....	113
Gráfico 25: Agência 02 (com outliers). Série completa.	140
Gráfico 26: Agência 02 (com outliers) – Parte superior representando a tendência e ciclo da série e na parte inferior a irregularidade.	140
Gráfico 27: Agência 02 (sem outliers). Série completa.....	141
Gráfico 28: Agência 02 (sem outliers) – Parte superior representando a tendência e ciclo da série e na parte inferior a irregularidade.	141
Gráfico 29: Agência 02 (com outliers) – Histograma referente às componentes tendência/ciclo (retângulo tracejado) e irregularidade (retângulo sólido).....	142
Gráfico 30: Agência 02 (sem outliers) – Histograma referente às componentes tendência/ciclo (retângulo tracejado) e irregularidade (retângulo sólido).....	143
Gráfico 31: Agência 06 (com outliers). Série completa.	143
Gráfico 32: Agência 06 (com outliers) – Parte superior representando a tendência e ciclo da série e na parte inferior a irregularidade.	144
Gráfico 33: Agência 06 (sem outliers). Série completa.....	144
Gráfico 34: Agência 06 (sem outliers) – Parte superior representando a tendência e ciclo da série e na parte inferior a irregularidade.	145
Gráfico 35: Agência 06 (com outliers) – Histograma referente às componentes tendência/ciclo (retângulo tracejado) e irregularidade (retângulo sólido).....	146
Gráfico 36: Agência 06 (sem outliers) – Histograma referente às componentes tendência/ciclo (retângulo tracejado) e irregularidade (retângulo sólido).....	146
Gráfico 37: Agência 13 (com outliers). Série completa.	147
Gráfico 38: Agência 13 (com outliers) – Parte superior representando a tendência e ciclo da série e na parte inferior a irregularidade.	147

Gráfico 39: Agência 13 (sem outliers). Série completa.....	148
Gráfico 40: Agência 13 (sem outliers) – Parte superior representando a tendência e ciclo da série e na parte inferior a irregularidade.	148
Gráfico 41: Agência 13 (com outliers) – Histograma referente às componentes tendência/ciclo (retângulo tracejado) e irregularidade (retângulo sólido).....	149
Gráfico 42: Agência 13 (sem outliers) – Histograma referente às componentes tendência/ciclo (retângulo tracejado) e irregularidade (retângulo sólido).....	150
Gráfico 43: Agência 02 – Gráfico dos dados reais (linha pontilhada) e previstos (linha contínua) (modelo Box e Jenkins).....	151
Gráfico 44: Agência 02 – Gráfico dos dados reais (linha pontilhada) e previstos (linha contínua) (modelo amortecimento exponencial).....	151
Gráfico 45: Agência 06 – Gráfico dos dados reais (linha pontilhada) e previstos (linha contínua) (modelo Box e Jenkins).....	152
Gráfico 46: Agência 06 – Gráfico dos dados reais (linha pontilhada) e previstos (linha contínua) (modelo amortecimento exponencial).....	152
Gráfico 47: Agência 13 – Gráfico dos dados reais (linha pontilhada) e previstos (linha contínua) (modelo Box e Jenkins).....	153
Gráfico 48: Agência 13 – Gráfico dos dados reais (linha pontilhada) e previstos (linha contínua) (modelo amortecimento exponencial).....	153
Gráfico 49: Representação da frequência experimental acumulada $S(x)$ e função teórica proposta $F(x)$	185
Gráfico 50: Agência 02 – Índices de autocorrelação por deslocamento.....	196
Gráfico 51: Agência 06 – Índices de autocorrelação por deslocamento.....	197
Gráfico 52: Agência 13 – Índices de autocorrelação por deslocamento.....	197

LISTA DOS QUADROS

Quadro 1: Quadro representativo do crescimento do meio circulante nacional.....	27
Quadro 2: Valores gastos com transporte de numerário. Fonte: Banco do Brasil S.A.	29
Quadro 3: Rio de Janeiro - agências da base de dados localizadas na capital.....	36
Quadro 4: Rio de Janeiro - agências da base de dados localizadas no interior.	36
Quadro 5: Características das agências com base no movimento e na oscilação. São consideradas pagadoras as agências cujos movimentos de saques foram iguais ou superiores a 60% dos dados da série, as receptoras são assim classificadas se tiveram depósitos em quantidade igual ou superior a 60% da série. As demais são classificadas como equilibrada. Já as oscilações foram consideradas baixas se o desvio padrão foi menor ou igual a 100, para o intervalo (100, 500] classificou-se como média e acima deste patamar consideramos uma agência como de alta oscilação.	38
Quadro 6: Valores estatísticos das agências.	40
Quadro 7: Correlação das agências. Na matriz triangular superior estão destacados as menores correlações (índice de correlação menor ou igual a 0,10).....	40
Quadro 8: Agência 02. Valores mínimos, máximos, médias e desvios padrões dos dias da semana.	48
Quadro 9: Agência 06. Valores mínimos, máximos, médias e desvios padrões dos dias da semana.	50
Quadro 10: Agência 13. Valores mínimos, máximos, médias e desvios padrões dos dias da semana.	50
Quadro 11: Agência 02. Valores mínimos, máximos, médias e desvios padrões de cada dia útil do mês.	53
Quadro 12: Agência 06. Valores mínimos, máximos, médias e desvios padrões de cada dia útil do mês.	54
Quadro 13: Agência 13. Valores mínimos, máximos, médias e desvios padrões de cada dia útil do mês.	55

Quadro 14: Análise de ciclicidade da série. Vê-se que no deslocamento de 1 posição há coincidência em 5 registros que estão sombreados.	71
Quadro 15: Agência 02 – Comparação dos dias com base completada com valor positivo.	72
Quadro 16: Agência 06 – Comparação dos dias com base completada com valor positivo. Não há informação nas áreas sombreadas em função de a janela de 5 posições haver tido o melhor desempenho.	72
Quadro 17: Agência 13 – Comparação dos dias com base completada com valor positivo.	72
Quadro 18: Agência 02 – Comparação dos dias com base completada com valor negativo.	73
Quadro 19: Agência 06 – Comparação dos dias com base completada com valor negativo.	73
Quadro 20: Agência 13 – Comparação dos dias com base completada com valor negativo.	73
Quadro 21: Agência 02 – Comparação dos dias com a base completa.	73
Quadro 22: Agência 06 – Comparação dos dias com a base completa.	74
Quadro 23: Agência 13 – Comparação dos dias com a base completa.	74
Quadro 24: Outliers encontrados nas séries de dados das agências 02, 06 e 13.	75
Quadro 25: Quadro resumo dos feriados nacionais no período sob estudo.	76
Quadro 26: Relação de feriados geradores de outliers para duas agências simultaneamente.	77
Quadro 27: Resumo da quantidade de outliers por agência. A coluna “P” indica a quantidade de outliers referentes a movimentos de pagamentos e a coluna “R” de recebimentos.	79
Quadro 28: Resumo da quantidade de outliers por agência e dia da semana. A coluna “P” indica a quantidade de outliers referentes a movimentos de pagamentos e a coluna “R” de recebimentos.	80
Quadro 29: Resumo da quantidade de outliers por agência e mês do ano. A coluna “P” indica a quantidade de outliers referentes a movimentos de pagamentos e a coluna “R” de recebimentos.	81
Quadro 30: Análise da relevância das entradas para as agências 02, 06 e 13.	123

Quadro 31: Exemplos de conjuntos tradicionais.	132
Quadro 32: Propriedades Fundamentais das operações sobre Conjuntos Tradicionais	133
Quadro 33: Operações e relações com conjuntos nebulosos.	134
Quadro 34: Agência 02. Estatística das componentes da série.	142
Quadro 35: Agência 06. Estatística das componentes da série.	145
Quadro 36: Agência 13. Estatística das componentes da série.	149
Quadro 37: Decomposição clássica das séries relativas às agências 02, 06 e 13	150
Quadro 38: Principais parâmetros adotados na construção das RNA.	159
Quadro 39: Principais parâmetros adotados na construção das RNA.	160
Quadro 40: Principais parâmetros na criação de redes para verificar a quantidade ideal de neurônios. Janela de 5.	162
Quadro 41: Principais parâmetros na criação de redes para verificar a quantidade ideal de neurônios. Janela de 10.	163
Quadro 42: Meses/anos cujos primeiros dias úteis do mês são coincidentes.	164
Quadro 43: Principais parâmetros na criação de redes com conjunto de treinamento somente dos meses cujo 1º. dia útil seja a segunda-feira.	165
Quadro 44: Principais parâmetros na criação de redes “Elman”.	166
Quadro 45: Acertos na direção dos dados previstos pela melhor reElman para d_t e d_{t-1}	166
Quadro 46: Principais parâmetros na criação de redes de MLP e Elman com uso da fórmula de normalização 1.	167
Quadro 47: Principais parâmetros das melhores redes de MLP e Elman usadas na primeira fase.	169
Quadro 48: Principais parâmetros das redes de MLP e Elman com função de propagação não linear na camada de saída.	170
Quadro 49: Principais parâmetros das redes de MLP com cálculo de performance através de intervalos.	172
Quadro 50: Principais parâmetros das redes de MLP com cálculo de performance através da divisão de 2 a 10 intervalos	173
Quadro 51: Principais parâmetros das redes de MLP e Elman no experimento 10 . . .	174
Quadro 52: Principais parâmetros das redes de MLP no experimento 11	175
Quadro 53: Principais parâmetros das redes de MLP no experimento 12.	176

Quadro 54: Distribuição de frequência de propriedade de automóvel por curso.	183
Quadro 55: Frequências observadas e esperadas.....	183
Quadro 56: Teste de Kolmogov-Sminorv – Valores sorteados.....	186
Quadro 57: Resumo dos procedimentos 1 a 5 do teste de Kolmogorov-Smirnov.....	187
Quadro 58: Seqüência hipotética.....	188
Quadro 59: Quantidade de dióxido de enxofre na atmosfera durante 44 dias sucessivos.....	92
Quadro 60: Seqüência aleatória criada por um gerador de números randômicos.	194
Quadro 61: Os 10 maiores índices de correlação para as séries sob estudo.....	196
Quadro 62: Teste qui-quadrado para significância de 0,05 e várias frequências.	199
Quadro 63: Teste de K-S com nível de significância 0,05.	201
Quadro 64: Teste de repetições de 2 observações, com significâncias 0,05 e 0,01....	202
Quadro 65: Teste de repetições acima e abaixo da mediana, com significâncias 0,05 e 0,01.....	203
Quadro 66: Teste de repetições para cima e para baixo, com significâncias 0,05 e 0,01.....	203
Quadro 67: Quadro comparativo dos modelos utilizados.....	209
Quadro 68: Ocorrências verificadas nas agências 02, 06 e 13 em função dos feriados do ano de 1998.....	211
Quadro 69: Ocorrências verificadas nas agências 02, 06 e 13 em função dos feriados do ano de 1999.....	212
Quadro 70: Ocorrências verificadas nas agências 02, 06 e 13 em função dos feriados do ano de 2000 e 2001, existentes até 31/03.	213
Quadro 71: Agência 02 – Distribuição mensal dos outliers.	214
Quadro 72: Agência 06 – Distribuição mensal dos outliers.	214
Quadro 73: Agência 13 – Distribuição mensal dos outliers.	214
Quadro 74: Agência 02 – outliers classificados por dia útil seguido da data. A coluna “semana” representa o dia da semana, segundo a seguinte correlação: 2 corresponde à segunda-feira, 3 à terça-feira e assim sucessivamente. Na coluna “mês” temos os dias do mês e, por último, na coluna “útil” o dia útil a que se refere.....	215

Quadro 75: Agência 06 – Outliers classificados por dia útil seguido da data. A coluna “semana” representa o dia da semana, segundo a seguinte correlação: 2 corresponde à segunda-feira, 3 à terça-feira e assim sucessivamente. Na coluna “mês” temos os dias do mês e, por último, na coluna “útil” o dia útil a que se refere.....	215
Quadro 76: Agência 13 – Outliers classificados por dia útil seguido da data. A coluna “semana” representa o dia da semana, segundo a seguinte correlação: 2 corresponde à segunda-feira, 3 à terça-feira e assim sucessivamente. Na coluna “mês” temos os dias do mês e, por último, na coluna “útil” o dia útil a que se refere.....	216
Quadro 77: Ag. 02 – Outliers classificados por dia da semana. A coluna “semana” representa o dia da semana, segundo a seguinte correlação: 2 corresponde à segunda-feira, 3 à terça-feira e assim sucessivamente. Na coluna “mês” têm-se os dias do mês e, por último, na coluna “útil” o a que se refere.	216
Quadro 78: Agência 06 – Outliers classificados por dia da semana seguido da data. A coluna “semana” representa o dia da semana, segundo a seguinte correlação: 2 corresponde à segunda-feira, 3 à terça-feira e assim sucessivamente. Na coluna “mês” têm-se os dias do mês e, por último, na coluna “útil” o a que se refere.	217
Quadro 79: Agência 13 – Outliers classificados por dia da semana seguido da data. A coluna “semana” representa o dia da semana, segundo a seguinte correlação: 2 corresponde à segunda-feira, 3 à terça-feira e assim sucessivamente. Na coluna “mês” têm-se os dias do mês e, por último, na coluna “útil” o a que se refere.	217
Quadro 80: Previsão da série da agência 02 (sem outliers) segundo os modelos de Box-Jenkins e amortecimento exponencial.	218
Quadro 81: Previsão da série da agência 06 (sem outliers) segundo os modelos de Box-Jenkins e amortecimento exponencial.	219
Quadro 82: Previsão da série da agência 13 (sem outliers) segundo os modelos de Box-Jenkins e amortecimento exponencial.	220
Quadro 83: Quadro resumo das melhores topologias nos testes de redes neurais utilizando 4 formas de normalizações lineares.....	221
Quadro 84: Resultados relativos aos treinamentos com as melhores redes.....	221
Quadro 85: Acertos na direção dos dados previstos pela melhor rede para d e $d - 1$.	221

Quadro 86: Resultados do treinamento de redes usando a técnica de CPD para a série de dados da agência 02. Em destaque têm-se o melhor resultado alcançado.....	222
Quadro 87: Resultados do treinamento de redes usando a técnica de CPD para a série de dados da agência 06. Em destaque têm-se os melhores resultados alcançados.....	222
Quadro 88: Resultados do treinamento de redes usando a técnica de CPD para a série de dados da agência 13. Em destaque têm-se os melhores resultados alcançados.....	223
Quadro 89: Resultados do treinamento de redes usando a técnica de CPD para a série de dados da agência 02. Em destaque têm-se o melhor resultado alcançado.....	223
Quadro 90: Resultados do treinamento de redes usando a técnica de CPD para a série de dados da agência 06. Em destaque têm-se os melhores resultados alcançados.....	224
Quadro 91: Resultados do treinamento de redes usando a técnica de CPD para a série de dados da agência 13. Em destaque têm-se o melhor resultado alcançado.....	224
Quadro 92: Agência 02 - Resultados do treinamento de redes Elman. Em destaque têm-se o melhor resultado alcançado.....	225
Quadro 93: Agência 06 - Resultados do treinamento de redes Elman. Em destaque têm-se o melhor resultado alcançado.....	225
Quadro 94: Agência 13 - Resultados do treinamento de redes Elman. Em destaque têm-se o melhor resultado alcançado.....	226
Quadro 95: Resumo dos melhores resultados no treinamento de redes MLP e Elman para as agências 02, 06 e 13, com 1 a 50 neurônios na primeira camada e utilizando 4 formas de normalizações não lineares.....	226
Quadro 96: Agência 02 - Treinamento das melhores redes de MLP e Elman usadas na primeira fase.....	227
Quadro 97: Agência 06 - Treinamento das melhores redes de MLP e Elman usadas na primeira fase.....	227
Quadro 98: Agência 13 - Treinamento das melhores redes de MLP e Elman usadas na primeira fase.....	228
Quadro 99: Resumo das melhores redes MLP e Elman treinadas para as agências 02 usando funções de propagação não lineares nas duas camadas.....	228

Quadro 100: Resumo das melhores redes MLP e Elman treinadas para as agências 06 usando funções de propagação não lineares nas duas camadas.....	228
Quadro 101: Resumo das melhores redes MLP e Elman treinadas para as agências 13 usando funções de propagação não lineares nas duas camadas.....	229
Quadro 102: Agência 02 – Performance da rede em 4 e 5 intervalos.	229
Quadro 103: Agência 02 – Performance da rede em 6 e 8 intervalos.	230
Quadro 104: Ag. 06 – Performance da rede em 4 e 5 intervalos.....	230
Quadro 105: Agência 06 – Performance da rede em 6 e 8 intervalos.	231
Quadro 106: Ag. 13 – Performance da rede em 4 e 5 intervalos.....	231
Quadro 107: Agência 13 – Performance da rede em 6 e 8 intervalos.	232
Quadro 108: Performance alcançada por intervalo.	232
Quadro 109: Ag. 02 – Performance binarização dos dados.....	233
Quadro 110: Ag. 06 – Performance binarização dos dados.....	233
Quadro 111: Ag. 13 – Performance binarização dos dados.....	234
Quadro 112: Ag. 02 – Performance binarização dos dados – experimento 11.....	235
Quadro 113: Ag. 06 – Performance binarização dos dados – experimento 11.....	236
Quadro 114: Ag. 13 – Performance binarização dos dados – experimento 11.....	237
Quadro 115: Agência 02 – Performance utilizando média móvel – experimento 12.238	
Quadro 116: Agência 06 – Performance utilizando média móvel – experimento 12.238	
Quadro 117: Agência 13 – Performance utilizando média móvel – experimento 12.238	
Quadro 118: Ag. 02 – Resultados da aplicação dos modelos fuzzy.....	239
Quadro 119: Ag. 06 – Resultados da aplicação dos modelos fuzzy.....	240
Quadro 120: Ag. 13 – Resultados da aplicação dos modelos fuzzy.....	241

ANEXOS

1. Análise dos Outliers	211
2. Tratamento dos <i>Outliers</i>	214
3. Modelos Estatísticos Utilizados	218
4. Modelos e Experimentos Realizados – Experimento 1	221
5. Modelos e Experimentos Realizados – Experimento 2	222
6. Modelos e Experimentos Realizados – Experimento 3	223
7. Modelos e Experimentos Realizados – Experimento 5	225
8. Modelos e Experimentos Realizados – Experimento 6	226
9. Modelos e Experimentos Realizados – Experimento 7	228
10. Modelos e Experimentos Realizados – Experimento 8	229
11. Modelos e Experimentos Realizados – Experimento 9	232
12. Modelos e Experimentos Realizados – Experimento 10	233
13. Modelos e Experimentos Realizados – Experimento 11	235
14. Modelos e Experimentos Realizados – Experimento 12	238
15. Modelo Nebuloso	239

SUMÁRIO

Agradecimentos	V
Resumo	VI
Abstract.....	VII
Lista das Figuras	VII
Lista dos Gráficos	X
Lista dos Quadros	XIII
Anexos	XX
Sumário.....	XXI
1. Introdução.....	27
1.1. Panorama Atual	27
1.2. O Modelo de Distribuição de numerário	30
1.3. Objetivos.....	31
1.4. Organização da Dissertação.....	32
2. Análise Preliminar do Problema.....	35
2.1. Descrição do Problema	35
2.2. Análise Subjetiva.....	36
2.2.1. Classificação das Agências.....	37
2.3. Estudo Comparativo do Comportamento das Agências	39
2.3.1. O tratamento das agências 02, 06 e 13	42
2.3.2. Análise Mensal dos Movimentos das Agências 02, 06 e 13	44
2.3.3. Análise Semanal dos Movimentos das Agências 02, 06 e 13	48
2.3.4. Análise Diária dos Movimentos das Agências 02, 06 e 13	52
2.4. Conclusões.....	56
2.5. Hipóteses	57
3. Ferramentas de Análise de Séries Temporais.....	59
3.1. Introdução.....	59
3.2. Ferramentas de Análise	59
3.2.1. Medidas de Tendência Central	60

Média Aritmética.....	60
Média Móvel	61
Mediana	61
Moda.....	61
3.2.2. Medidas de Dispersão	62
Variância	62
Desvio padrão.....	62
3.2.3. Coeficiente de Correlação	63
3.3. Transformação de Dados	64
3.4. Valores Inexistentes (“Missing Values”).....	66
3.5. valores espúrios (“Outliers”)	67
3.6. Análise do Capítulo	68
4. Análise das Séries em Estudo.....	70
4.1. Tratamento dos Dados	70
4.2. Tratamento dos outliers	75
4.3. Normalização da Base de Dados	83
5. Ferramentas de Modelagem Para Previsão.....	87
5.1. Ferramentas Clássicas.....	87
5.2. Séries Temporais	88
5.3. Amortecimento Exponencial	89
5.3.1. Amortecimento Exponencial Simples	89
Método de Holt.....	92
Método de Holt-Winters.....	93
Modelo Multiplicativo de Holt-Winters.....	94
Modelo Aditivo de Holt-Winters	95
5.4. Modelo ARIMA - AutoRegressive Integrated Moving Average	96
5.5. Ferramentas Não Clássicas	98
5.6. Redes Neurais Artificiais.....	99
5.6.1. O Modelo Multilayer Perceptron.....	101
5.6.2. Treinamento e Operação de Redes Neurais do Tipo MLP	102
5.6.3. O Modelo “Elman”	106
5.6.4. Redes com Atraso no Tempo	108

5.7. Estratégia, Treinamento e Construção de Redes Neurais.....	109
5.8. Escolha da Estrutura da Rede (Arquitetura).....	111
5.8.1. Quantidade de Camadas.....	113
5.8.2. Quantidade de Neurônios.....	114
5.8.3. Conexões.....	115
5.8.4. Funções de Ativação e de Propagação.....	115
5.9. Dinâmica ou Treinamento da Rede.....	118
5.10. Redução de Dimensionalidade.....	119
5.10.1. Análise de Componentes Principais (PCA).....	119
5.10.2. Relevância das Entradas.....	122
5.10.3. Componente Principal de Discriminação – CPD.....	124
5.11. Regras de Aprendizado.....	125
5.11.1. Taxa de Aprendizado.....	126
5.11.2. Momento.....	128
5.12. Lógica Nebulosa.....	130
5.12.1. Teoria dos Conjuntos Tradicionais.....	131
5.12.2. Teoria dos Conjuntos Nebulosos.....	133
Corte α	134
Conjunto de Níveis.....	135
Suporte.....	135
Altura.....	135
Normalização.....	135
5.12.3. Sistemas Nebulosos.....	136
5.12.4. Regras Nebulosas.....	136
5.13. Conclusões do Capítulo.....	137
6. Abordagem Clássica das Séries em estudo.....	139
6.1. Componentes de Uma Série Temporal Segundo a Abordagem Clássica.....	139
6.2. Decomposição Gráfica das Componentes.....	139
6.3. Modelos Estatísticos Utilizados.....	150
6.4. Conclusões.....	154
7. Abordagem Não Clássica das Séries em estudo.....	155
7.1. Modelos e Experimentos Realizados.....	155

7.1.1. Pré-processamento dos Dados.....	156
7.1.2. Redes Neurais Artificiais.....	157
7.2. Experimento 1.....	158
Rede MLP com 3 Camadas e Normalização Linear	158
7.3. Experimento 2.....	161
Componente Principal de Discriminação – Janela de 5 elementos	161
7.4. Experimento 3.....	162
Componente Principal de Discriminação – Janela de 10 elementos	162
7.5. Experimento 4.....	163
Conjunto de Treinamento com Início de Mês Coincidente.....	163
7.6. Experimento 5.....	165
Redes “Elman” – Normalização Linear	165
7.7. Experimento 6.....	167
Redes MLP e “Elman” – Normalização Não Linear.....	167
7.8. Experimento 7.....	169
Redes MLP e Elman – Funções de propagação não lineares	169
7.9. Experimento 8.....	171
Redes MLP – Verificação da performance da rede através de intervalos.....	171
7.10. Experimento 9.....	172
Redes MLP com quantização dos dados	172
7.11. Experimento 10.....	173
Redes MLP e “Elman” com binarização.....	173
7.12. Experimento 11.....	174
Redes MLP com binarização.....	174
7.13. Experimento 12.....	175
Redes MLP com média móvel	175
7.14. Modelo Nebuloso	177
7.15. Conclusão	178
8. Estudo da Aleatoriedade da Série.....	179
8.1. Amostragem Aleatória.....	179
8.2. Método da Autocorrelação	180
8.3. Método do Qui-Quadrado.....	182

8.4. Método de Kolmogorov-Smirnov	184
8.5. Método de Corridas	187
8.5.1. Corrida.....	188
8.5.2. Análise das Repetições	188
8.5.3. Teste de repetições de dois tipos de observações.....	189
8.5.4. Teste de repetições acima e abaixo da mediana	191
8.5.5. Teste de repetições acima e abaixo em relação ao anterior.....	193
8.6. Resultados.....	195
8.6.1. Teste da Autocorrelação	195
8.6.2. Teste do qui-quadrado	198
8.6.3. Teste de Kolmogorov-Smirnov	200
8.6.4. Teste de repetições de 2 tipos de observações.....	202
8.6.5. Teste de repetições acima e abaixo da mediana	202
8.6.6. Teste de repetições acima e abaixo em relação ao anterior.....	203
8.7. Conclusões	203
9. Conclusão	206
9.1. Introdução	206
9.2. Principais Dificuldades Encontradas	207
9.3. Modelos Aplicados	208
9.3.1. Modelo de amortecimento exponencial e ARIMA	208
9.3.2. Modelo Neural.....	209
9.3.3. Modelo Nebuloso	209
9.3.4. Comparações Entre os Diversos Modelos.....	209
9.4. Trabalhos Futuros	209
Anexo	211
1. Análise dos Outliers.....	211
2. Tratamento dos <i>Outliers</i>	214
3. Modelos Estatísticos Utilizados.....	218
4. Modelos e Experimentos Realizados – Experimento 1	221
5. Modelos e Experimentos Realizados – Experimento 2	222
6. Modelos e Experimentos Realizados – Experimento 3	223
7. Modelos e Experimentos Realizados – Experimento 5	225

8. Modelos e Experimentos Realizados – Experimento 6	226
9. Modelos e Experimentos Realizados – Experimento 7	228
10. Modelos e Experimentos Realizados – Experimento 8	229
11. Modelos e Experimentos Realizados – Experimento 9	232
12. Modelos e Experimentos Realizados – Experimento 10	233
13. Modelos e Experimentos Realizados – Experimento 11	235
14. Modelos e Experimentos Realizados – Experimento 12	238
15. Modelo Nebuloso	239
10. Referências	242

1. INTRODUÇÃO

Este capítulo inicia com uma visão geral a respeito das questões que envolvem a movimentação de dinheiro entre agências de bancos. Em seguida discorre sobre o modelo de distribuição adotado por um grande banco brasileiro para, na seção seguinte, apresentar os objetivos do trabalho. Termina descrevendo a organização da dissertação.

1.1. PANORAMA ATUAL

Desde a criação do papel-moeda (numerário no meio técnico) como elemento de troca entre as pessoas, a cobiça fez com que ele fosse um bem extremamente desejado, tendo em vista o alto valor intrínseco, a facilidade de manuseio e liquidez. Dessa forma, a sua manipulação, transporte e controle necessitam de logística especial e, em razão disso, o seu custo operacional é muito alto.

Mesmo com a crescente utilização por meios eletrônicos de pagamento que permitiram infinidades de transações para transferência de recursos via sistema, como os cartões de crédito, internet *banking*, cartões de débito, etc., o indicador macroeconômico que mede a quantidade de papel-moeda em circulação – Meio Circulante Nacional – registra elevações crescentes a cada ano, conforme quadro abaixo (BANCO CENTRAL DO BRASIL, 2003). A idéia de que com o avanço tecnológico o “dinheiro de plástico” e as transferências *on-line* reduziriam o montante do meio circulante não vem se concretizando mas, ao contrário, o numerário em circulação vem aumentando constantemente.

Quadro 1: Quadro representativo do crescimento do meio circulante nacional

Posição em	R\$ bilhões
31/12/1994	10,046
31/12/1995	13,770
31/12/1996	17,188
31,12/1997	20,251
31/12/1998	24,167
31/12/1999	29,838
31/12/2000	32,633
31/12/2001	37,669
31/12/2002	49,931

As instituições bancárias de todo o mundo realizam, diariamente, milhares de viagens para entrega e recolhimento de dinheiro junto às suas filiais. O montante despendido nestas operações é muito elevado fazendo com que o planejamento referente às fases de distribuição do numerário seja uma importante etapa para minimizar os custos inerentes a este procedimento.

Entretanto, trata-se de um exercício extremamente complexo, pois diversas variáveis fazem parte do ambiente a ser analisado. Como exemplo, podemos citar o dia da semana (segunda-feira etc.), o dia do mês (primeiro dia útil etc.), o mês do ano (janeiro, fevereiro etc.), se o dia é véspera ou logo após um feriado, a localização da agência (cidade turística, industrial etc.) e outras mais. Até o tempo tem influência no movimento bancário. A distância entre as agências também é um fator importante nesta equação pois a remessa, às vezes, necessita ser despachada com vários dias de antecedência. Assim, a previsão deve ser vista não somente em $d + 1$ mas, em alguns casos, também em $d + 2$, $d + 3$, onde d corresponde ao dia vigente.

A otimização dos recursos envolvidos no trato do dinheiro é meta de todos os bancos pois a melhoria nas etapas de tratamento do numerário reduz despesas operacionais. Dentre todas, a mais sensível e, portanto, a mais custosa, é o seu transporte, pois ele envolve um aparato de controles e segurança muito grande. Não raro, o carro-forte necessita pernoitar fora de sua base, consumindo despesas com estacionamento, vigilância ininterrupta, etc. Além disso, a remessa/alívio solicitada em intervalos menores que 24 horas, acarreta uma penalidade imposta pelas empresas transportadoras de 100% do valor original, onerando, ainda mais, o custo do transporte.

Atualmente o custo com transporte de numerário, entre as dependências de uma instituição bancária, vem crescendo (quadro 2) em função de estratégias operacionais que visam manter em estoque o mínimo de dinheiro possível nas agências, a fim de que na eventualidade de ocorrência de sinistros os prejuízos sejam minimizados. Entretanto, com o receio de que possa ocorrer falta de recursos para atendimento à clientela, gerando sérios problemas à imagem da Instituição, os administradores das dependências mantêm margem ociosa de numerário que permanece constantemente nos cofres da agência.

Quadro 2: Valores gastos com transporte de numerário. Fonte: Banco do Brasil S.A.

Ano	R\$ milhões
-----	-------------

1999	104,417
2000	147,703
2001	181,716
2002	225,079

Apesar de conhecer a situação, os responsáveis pelo planejamento estratégico da empresa ficam preocupados em forçarem os administradores das agências a reduzirem a margem ociosa, tendo em vista não possuírem uma previsão confiável das reais necessidades de numerário para o atendimento dos compromissos do Banco com seus clientes.

A otimização dos recursos a permanecerem nos cofres da dependência assim como a redução do número de viagens de suprimento ou recolhimento de dinheiro nas agências é fator preponderante para se minimizar os custos operacionais do sistema e vem sendo buscada pelos analistas das empresas.

Todas as agências possuem um limite máximo de numerário a permanecer em seus cofres, estabelecido por órgão central do Banco, mediante análise do comportamento da dependência. Baseado neste parâmetro o administrador da agência solicita suprimento ou recolhe valores para a agência supridora. Acontece que este parâmetro sofre influência do modo de administrar do gerente pois um gerente conservador pode trabalhar com um mínimo de dinheiro no cofre realizando diversas viagens de recolhimentos e suprimentos; já outro administrador pode ter uma característica mais liberal, visando economia de recursos com o transporte, e, neste caso, realizar menos viagens. Têm também os casos das agências metropolitanas que em razão de terem um custo de transporte menor são estimuladas a realizarem viagens de intercâmbio com maior frequência que aquelas situadas em localidades cujos custos são mais elevados.

A fim de excluir a influência da componente acima mencionada na dinâmica do processo, ou seja, a interferência do administrador da agência, optou-se por modelar o problema com base somente na variação diária do movimento da agência decorrente de saques e depósitos feitos pelos clientes. Esta estratégia permite estudar tão-somente os padrões de comportamento dos clientes, foco desta dissertação e, para tanto, foi adotada a seguinte equação:

$\text{Variação} = \text{Saldo Final} + \text{Alívio} - \text{Reforço} - \text{Saldo Inicial}$	Eq. (1)
--	---------

Os valores diários referentes à variação calculada na equação acima são as componentes da série temporal. Cabe lembrar que variações negativas significam que naquele dia houve saque de recursos da dependência, ou seja, o público retirou mais dinheiro do que depositou. No caso inverso, ou seja, variações positivas, representam a entrada de recursos na agência.

1.2. O MODELO DE DISTRIBUIÇÃO DE NUMERÁRIO

O Banco que vimos investigando trabalha com uma rede de distribuição em forma de árvore. O numerário fabricado na Casa da Moeda é distribuído para 10 (dez) delegacias localizadas nas principais capitais do País. De lá é encaminhado para agências de maior porte e, vai sendo levado para localidades de menor porte até se alcançar a agência-folha.

Uma dependência encarregada de fornecer numerário para outras é denominada de “supridora”. Aquelas que tão-somente recebem dinheiro para seu uso, sem responsabilidade de suprir outras agências com numerário, chama-se “caudatária”. Este modelo é conhecido como “Rede de Intercâmbio de Numerário”.

No total este Banco possui aproximadamente 8.500 pontos de fornecimento de dinheiro, aí computado agências, *cash-dispensers*, postos de serviços e estandes em feiras e eventos e realiza, mensalmente, algo em torno de 35.000 viagens de numerário, sendo que algumas, em virtude da extensão de nosso território e da capilaridade da rede de agências, são feitas utilizando-se avião no transporte. Os custos para se realizar o transporte deste dinheiro oscila na faixa de R\$ 19 milhões/mês.

No processo de intermediação financeira em que estão inseridos os bancos comerciais, 3 situações podem ocorrer: a primeira acontece quando os recursos existentes na dependência não são suficientes para fazer face aos compromissos assumidos, necessitando, portanto, de reforço de numerário em seus cofres; a segunda quando muitos depósitos são realizados pelo público e, então, surge a necessidade de se remanejar o excedente, por questões de segurança, para outra filial e, por último, quando o volume de numerário existente encontra-se equilibrado com os compromissos a curto prazo. Nas duas primeiras situações a figura do transporte de numerário faz-se presente, a fim de se deslocar os recursos de um ponto para outro.

1.3. OBJETIVOS

O principal objetivo deste trabalho é o de conceber um modelo computacional que forneça previsões da necessidade de numerário de agências bancárias de uma determinada instituição, num horizonte de $d + 1$, onde d corresponde ao dia corrente. Para tanto se avaliou o uso de modelos econométricos, neurais e nebulosos.

Os modelos clássicos, amplamente utilizados em economia, foram utilizados a fim de se estudar o comportamento da série.. Os modelos não clássicos, de uso mais recente, foram aplicados visando avaliar a sua performance e servir como alternativa à abordagem tradicional.

O modelo estatístico mais em voga atualmente para tratamento de séries temporais financeiras é o ARIMA (*AutoRegressive Integrated Moving Average Models*) e, por esta razão, abordou-se os seus conceitos e se promoveu estudos comparativos com os resultados alcançados através do uso das demais técnicas.

As redes neurais artificiais (RNA) vêm se mostrando muito eficazes na previsão de situações baseadas em séries históricas. As RNA podem ser tratadas como aproximadores de funções matemáticas complexas e, nestes casos, vêm obtendo, geralmente, um desempenho superior aos modelos estatísticos convencionais (ZHANG; PATUWO; HU, 1988, p. 36).

Pesquisa-se o uso de redes neurais que sejam eficientes para as diversas características das agências no País, ou seja, respondam de forma objetiva e correta os questionamentos relativos à necessidade de papel-moeda nas dependências localizadas em regiões urbanas, turísticas ou outra qualquer.

A fim de atender aos requisitos acima, estudos sobre a topologia de várias redes neurais foram implementados, buscando-se alcançar o melhor desempenho possível. Trabalhou-se primordialmente com redes do tipo MLP – MultiLayer Perceptron e “Elman” por serem consideradas as mais eficazes no tratamento de séries temporais (BRAGA; CARVALHO; LUDEMIR, 1998, p. 198) (ZHANG; PATUWO; HU, 1988, p. 37).

Apesar de a utilização de redes neurais artificiais no tratamento de séries temporais virem se mostrando eficientes, os modelos estatísticos ainda são de larga utilização nos trabalhos envolvendo séries financeiras. Atentos a isto, efetuou-se

avaliação dos resultados obtidos comparando-os com as previsões realizadas com o uso de técnicas de estatística aplicadas a séries temporais.

Por fim, avaliou-se o uso de modelo nebuloso. A Lógica Nebulosa (*fuzzy* ou difusa) é uma forma elegante e altamente eficaz de solucionar problemas, pois sistemas construídos com base neste paradigma são mais compactos e baratos que aqueles que utilizam soluções convencionais. Além disso, esses sistemas possuem a capacidade de processar dados inexatos ou parcialmente corretos. De fato, os sistemas difusos vêm sendo utilizados em diversos tipos de aplicações comerciais (GIZELLE, 2000, p.34).

A etapa de pré-processamento dos dados também foi foco deste trabalho, pois é de fundamental importância para o sucesso da pesquisa. O estudo dos dados, seu comportamento, a influência de feriados nos dias que o antecedem e o precedem, a ausência de valores na série, a relação entre dias de semanas equivalentes assim como dias de meses ou meses do ano foram alguns dos tópicos analisados e exaustivamente avaliados.

Em função da grande variância¹ verificada na série, vários experimentos foram realizados visando buscar a normalização mais eficiente para se aplicar nos dados. Assim, estudos com normalizações lineares e não lineares foram promovidos buscando a obtenção de uma resposta satisfatória dos modelos.

Obteve-se para esta pesquisa uma série com 811 registros, referente ao período de 02/01/1998 a 31/03/2001 (39 meses) de 15 agências localizadas no estado do Rio de Janeiro. Posteriormente acrescentou-se à base original mais 259 registros relativos ao período de 01/04/2001 à 31/03/2002. Com isto, ficou-se com um total de 1.059 registros abrangendo o período de 02/01/1998 à 28/03/2002 (51 meses). Entretanto, este último conjunto de dados só foi disponibilizado quando muito do trabalho já havia sido realizado. Desta forma, somente serão abordados os experimentos com a massa total de registros nos últimos testes.

1.4. ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação trata no capítulo 2 da análise preliminar do problema. Assim, no seu transcorrer, é feita avaliação das variáveis envolvidas e suas influências,

¹ Medida de dispersão que representa a proximidade entre os valores de um conjunto de dados.

de como se comportam as agências dependendo da localização geográfica em que estão inseridas e quais os fatores externos responsáveis pelas oscilações verificadas na série. Ainda no capítulo 2 foram realizadas análises referentes à forma de classificação das agências, buscando-se agrupá-las em conjuntos com características similares.

O capítulo 3 corresponde à conceituação formal das técnicas utilizadas na modelagem dos dados. Nele são apresentados os conceitos referentes às medidas de tendência, dispersão e coeficiente de correlação. Aborda-se os conceitos de séries temporais, formas de previsão, etc. Discorre-se, ainda, sobre as técnicas para separação das componentes envolvidas em uma série. As métricas de avaliação de desempenho, as formas de transformação de dados e o tratamento para os valores inexistentes e dados espúrios, conhecidos como *outliers*, também fazem parte do capítulo 3.

No capítulo 4 encontram-se os estudos promovidos durante a fase de pré-processamento dos dados. Apresentam-se os resultados alcançados através do uso de várias normalizações e a escolha da que melhor se adequou ao problema. Têm-se, também, o estudo dos *outliers* e uma abordagem sobre o efeito dos valores inexistentes na série, conhecidos como *missing values*.

As ferramentas de modelagem para previsão de séries temporais, aí considerando as abordagens clássicas (modelo econométrico) e não clássicas (redes neurais e lógica “fuzzy”), estão inseridas no capítulo 5. Nesta seção, discorre-se, também, sobre os conceitos dos diversos modelos. O desenvolvimento das equações econométricas, o paradigma neural, desde o mapeamento do neurônio biológico realizado no século XIX até a atualidade, as arquiteturas de redes utilizadas nesta dissertação e as técnicas implementadas para otimizar a rede e melhorar seu desempenho estão descritas nesta seção.

Dando seqüência ao trabalho, o capítulo 6 apresenta os resultados alcançados na previsão da necessidade de numerário das agências utilizando-se das técnicas da estatística clássica (amortecimento exponencial e Box e Jenkins).

A abordagem não clássica (redes neurais artificiais e lógica nebulosa) está compilada no capítulo 7.

No capítulo 8 fez-se um estudo sobre aleatoriedade de dados passando pela conceituação, apresentação de alguns métodos utilizados e, por fim, os resultados encontrados.

As conclusões do trabalho, assim como propostas para continuidade e melhoria desta dissertação são encontradas no capítulo 9.

2. ANÁLISE PRELIMINAR DO PROBLEMA

Este capítulo define, detalhadamente, o problema sob estudo. Faz uma análise preliminar e subjetiva das características das agências do banco. Termina discorrendo sobre as hipóteses levantadas e as conclusões obtidas.

2.1. DESCRIÇÃO DO PROBLEMA

Para estudar o problema referente a previsão da necessidade de numerário nas agências, obteve-se um conjunto de dados referente a 15 dependências situadas no estado do Rio de Janeiro. A escolha foi feita procurando-se obter pelo menos uma agência representativa de cada região do estado. Assim, buscou-se selecionar agências da capital e do interior, de região turística (serra, mar), localizadas em áreas industriais, subúrbios etc.

Dentre as agências localizadas na capital, foram definidas para efeito desta dissertação 4 categorias; já as do interior do estados foram classificadas em 3 categorias, conforme abaixo:

- Agências localizadas em capitais:
 - centro financeiro/comercial do município (centro);
 - áreas nobres da cidade (classe A);
 - áreas de classe média (classe B); e
 - áreas proletárias (classe C).
- Agências localizadas no interior:
 - cidades agropecuárias;
 - cidades turísticas; e
 - cidades industriais.
 -

Obviamente existem cidades que reúnem 2 ou mais das características comentadas e, neste caso, foi adotada aquela mais representativa da localidade.

A base de dados inicial compreende o período de 02/01/1998 à 31/03/2001 (39 meses), totalizando 811 registros para cada agência e o registro possui 5 variáveis:

data, saldo inicial, numerário encaminhado para outra agência (alívio), numerário recebido de outra agência (reforço) e saldo final. Posteriormente, o banco complementou a série fornecendo mais 249 registros, compreendendo o período de 03/04/2001 à 28/03/2002, perfazendo um total de 1.059 registros.

Por motivos de segurança, as agências sob estudo serão designadas somente por números, a fim de não se poder determinar a qual agência pertence as características apresentadas nesta dissertação. Entretanto, em alguns momentos, o nome da localidade é citado para facilitar o entendimento e melhor ilustrar o conceito que está sendo apresentado.

- Baseado nos parâmetros descritos acima, as agências disponíveis para estudo foram classificadas da seguinte forma:
- Rio de Janeiro – Capital:

Quadro 3: Rio de Janeiro - agências da base de dados localizadas na capital.

Centro	A	B	C
	Agência 05		Agência 01
	Agência 10		Agência 03

- Rio de Janeiro – Interior:

Quadro 4: Rio de Janeiro - agências da base de dados localizadas no interior.

Agropecuária	Industrial	Turística
Agência 07	Agência 04	Agência 02
Agência 09	Agência 08	Agência 06
Agência 13	Agência 12	Agência 11
Agência 14		
Agência 15		

2.2. ANÁLISE SUBJETIVA

É extremamente complexo analisar grupos de dependências pois cada cidade possui características próprias que dificultam definir seu comportamento. Podemos citar alguns exemplos: a cidade de Campos dos Goytacazes (RJ) é um importante centro de serviços e exerce uma forte influência nas cidades da região. Assim, nos finais de semana muitos habitantes de outras localidades vão à Campos fazer compras, tratamentos médicos etc. acarretando um fluxo de dinheiro para aquela região. Entretanto, este numerário será depositado nas agências localizadas nas áreas comerciais, ou seja, as dependências existentes nos bairros residenciais provavelmente não terão movimento em função deste fato.

O mesmo fenômeno de fluxo de dinheiro descrito para Campos dos Goytacazes (RJ) pode ocorrer em algumas agências localizadas nas capitais, como, por exemplo, a agência de Copacabana – Rio de Janeiro (RJ) onde o fato de receber muitos turistas faz com que o movimento das agências situadas naquele bairro aumente. O mesmo não se pode dizer para a agência de Bangu – Rio de Janeiro (RJ), pois se trata de região onde o movimento deve permanecer constante ao longo do ano pois nenhum comportamento similar aos descritos acima é verificado neste bairro do Rio de Janeiro. Baseado nestas considerações adotar uma análise por município não seria ideal. Haverá a necessidade de se estudar agência por agência, mesmo para aquelas situadas na mesma cidade.

Outro fator que deve estar presente é a questão da sazonalidade. Agências situadas em regiões praianas, a exemplo de Araruama (RJ), Cabo Frio (RJ) e outras, sofrem um considerável aumento populacional nos meses de verão, ocorrendo, também, um crescimento do movimento de dinheiro. Localidades rurais sofrem a influência das safras agrícolas, época em que se realiza a comercialização dos produtos com a conseqüente injeção de recursos na economia local. Possivelmente as agências situadas em áreas industriais devem ter um comportamento perene durante o ano, sem oscilações que caracterizem sazonalidade.

Apesar das dificuldades em se classificar as agências por características regionais, pode-se adotar como parâmetro o movimento das agências, ou seja, se a dependência possui características pagadora, recebedora ou equilibrada. Este, apesar de ser o critério mais aceito atualmente, ainda possui alguns inconvenientes neste tipo de análise, pois se o ano for considerado como parâmetro a dependência pode se comportar de uma forma durante um período (verão, por exemplo) e de outra durante outro período. Tem-se, ainda, a dificuldade em se definir a característica da agência durante o ano, ou seja, qual o período que seria ideal para considerá-la como uma agência pagadora: 3 meses, 6 meses ou mais? E se fosse pagadora durante 9 meses e fortemente recebedora durante os demais, qual a classificação que se daria?

2.2.1. Classificação das Agências

Como primeira proposição para classificação das dependências considerou-se uma agência como pagadora aquela em cujo movimento ocorreram saques em mais

de 60% dos dias sob estudo, já as receptoras são aquelas em que o movimento de depósitos correspondeu a mais de 60% de toda a série; e, por fim, o restante foi considerado como agências equilibradas.

Outra análise preliminar realizada diz respeito às oscilações nos movimentos. Adotou-se como parâmetro para classificar a agência com baixa oscilação aquela cujo desvio padrão (DP) seja de até R\$ 100 mil, àquelas situadas no intervalo (R\$ 100 mil, R\$ 500 mil] foram classificadas como de média oscilação e as com desvio padrão superior a R\$ 500 mil foram consideradas de alta oscilação. Esta classificação, assim como a do parágrafo anterior que trata do movimento de numerário realizado pelas dependências, pode ser vista no quadro a seguir.

Quadro 5: Características das agências com base no movimento e na oscilação. São consideradas pagadoras as agências cujos movimentos de saques foram iguais ou superiores a 60% dos dados da série, as receptoras são assim classificadas se tiveram depósitos em quantidade igual ou superior a 60% da série. As demais são classificadas como equilibrada. Já as oscilações foram consideradas baixas se o desvio padrão foi menor ou igual a 100, para o intervalo (100, 500] classificou-se como média e acima deste patamar consideramos uma agência como de alta oscilação.

Agências	Classificação	Movimento			Oscilação	
		Característica	Depósitos (%)	Saques (%)	Classificação	Desvio padrão
Ag. 01	Classe C	Equilibrada	56,10	43,90	Alta	723,73
Ag. 02	Turística	Equilibrada	42,77	56,23	Média	177,57
Ag. 03	Classe C	Pagadora	27,13	72,87	Média	327,91
Ag. 04	Industrial	Equilibrada	50,18	49,82	Média	423,56
Ag. 05	Classe A	Equilibrada	48,83	51,17	Média	149,12
Ag. 06	Turística	Recebedora	69,17	30,83	Alta	576,30
Ag. 07	Agropecuária	Equilibrada	43,40	56,60	Baixa	57,90
Ag. 08	Industrial	Pagadora	34,03	65,97	Alta	1.186,41
Ag. 09	Agropecuária	Recebedora	77,68	22,32	Baixa	67,12
Ag. 10	Classe A	Pagadora	16,52	83,48	Média	192,15
Ag. 11	Turística	Equilibrada	50,06	49,94	Alta	1.060,00
Ag. 12	Industrial	Pagadora	19,11	80,89	Média	340,15
Ag. 13	Agropecuária	Recebedora	70,65	29,35	Baixa	66,90
Ag. 14	Agropecuária	Equilibrada	43,65	56,35	Média	185,28
Ag. 15	Agropecuária	Pagadora	68,06	31,94	Baixa	72,51

Sob a ótica do movimento das agências verifica-se a existência de características diferentes, relativamente à localização geográfica das mesmas, ou seja, agências em áreas predominantemente agropecuárias têm características recebedora (agências 09 e 13), equilibrada (agências 07 e 14) e pagadora (agência 15). Em localidades turísticas temos agências equilibradas (agências 02 e 11) e do tipo recebedora (agência 06). Isto induz a se acreditar que o porte da agência, a atividade econômica do município ou, até mesmo, o fato de estar situada em uma região mais rica não caracteriza o tipo de movimento (equilibrado, pagador ou recebedor).

Entende-se que a característica do movimento está mais voltada para a atuação da administração da agência do que a localidade em que ela está inserida. Agências que possuem uma forte carteira de clientes comerciantes, industriais e outros tipos de atividades que gerem receitas devem ser, evidentemente, recebedoras. Já as situadas em locais cuja base de clientes são pessoas físicas, o movimento maior será de pagamentos. Por fim, aquelas onde há um equilíbrio na carteira de clientes, abrangendo não só pessoas físicas mas também o comércio, profissionais liberais e outros provavelmente possuem um movimento estável com relação aos saques e depósitos efetuados.

Analisando os grupos de agências baseado nas oscilações ocorridas no movimento vemos que há uma relação entre o comportamento dos desvios padrões das agências com as características geográficas. No grupo de oscilação baixa todas as agências ali listadas estão classificadas como sendo agropecuárias.

No grupo de oscilação média encontram-se agências de porte pequeno, localizadas em cidades sem apelo turístico, exceto a agência 02, ou de comércio. Têm-se ainda três das quatro agências do município do Rio de Janeiro corroborando a afirmação anterior de que o fato das viagens de numerário serem mais baratas facilitam e estimulam o intercâmbio de numerário com maior frequência.

Por fim, no grupo de agências com alta oscilação se encontram as três maiores agências de todo o conjunto – agências 06, 08 e 11. São agências localizadas em cidades de porte médio que sofrem influência, também, de uma grande população visitante, oriunda ou da microrregião em que exerce influência ou de turistas.

2.3. ESTUDO COMPARATIVO DO COMPORTAMENTO DAS AGÊNCIAS

Várias análises foram desenvolvidas visando conhecer melhor os dados existentes. Iniciando-se com o cálculo das médias, desvios padrões, valores máximos e mínimos da série e correlações, entre outros, foi-se, por etapas, avaliando as características dos dados. Esta é uma importante fase no estudo a ser efetuada, a fim de se ter um sólido conhecimento dos dados existentes para a futura modelagem. Baseado nestas análises promoveu-se estudos da melhor forma de normalização assim como o tratamento para os registros espúrios – *outliers* – que foram encontrados. A seguir têm-se os primeiros resultados estatísticos obtidos nas séries das agências.

Quadro 6: Valores estatísticos das agências.

Continue...	Média	Valor Mínimo	Valor Máximo	Variância	Desvio Padrão
AG_01	261.766	-796.81	5542.890	523788.	723.732
AG_02	-14.417	-590.20	1035.590	31531.	177.569
AG_03	-185.337	-1965.77	1208.350	107526.	327.911
AG_04	-99.318	-2519.12	1069.500	179402.	423.558
AG_05	-14.874	-1790.94	1599.570	22238.	149.123
AG_06	273.297	-1470.37	4227.610	332126.	576.304
AG_07	-10.595	-287.62	213.670	3352.	57.900
AG_08	-63.438	-5300.59	5066.570	1407568.	1186.410
AG_09	39.257	-149.75	549.900	4505.	67.123
AG_10	-200.437	-1115.40	307.970	36923.	192.154
AG_11	285.010	-2391.07	4153.110	1123609.	1060.004
AG_12	-278.661	-2037.97	991.860	115701.	340.148
AG_13	20.629	-265.84	919.400	4475.	66.898
AG_14	-52.009	-739.87	450.370	34330.	185.285
AG_15	-32.512	-500.00	550.000	5258.	72.514

Vê-se que a maioria das agências possui desvio padrão elevado, caracterizado pelas oscilações no movimento. Verifica-se, também, que apesar de a média situar-se em valores relativamente baixos o mesmo não se pode dizer em relação aos pontos mínimos e máximos, onde existem grandes variações. Estes pontos, apesar de não serem erros, pois foram conferidos, podem ser considerados elementos espúrios e, provavelmente, dificultarão o mapeamento a ser realizado pela rede.

Dando continuidade à análise estatística, foram calculados os coeficientes de correlações das dependências com o objetivo de verificar a existência de padrões de similaridade entre duas ou mais filiais.

Quadro 7: Correlação das agências. Na matriz triangular superior estão destacados as menores correlações (índice de correlação menor ou igual a 0,10).

	Ag01	Ag02	Ag03	Ag04	Ag05	Ag06	Ag07	Ag08	Ag09	Ag10	Ag11	Ag12	Ag13	Ag14	Ag15
Ag01	I	0,47	0,34	0,26	0,30	0,19	0,35	0,20	0,52	0,17	0,30	0,34	0,32	0,19	0,22
Ag02	0,47	I	0,51	0,51	0,33	0,37	0,48	0,42	0,54	0,29	0,53	0,59	0,41	0,43	0,49
Ag03	0,34	0,51	I	0,47	0,44	0,19	0,28	0,29	0,23	0,67	0,28	0,58	0,25	0,38	0,44
Ag04	0,26	0,51	0,47	I	0,25	0,45	0,31	0,55	0,29	0,31	0,62	0,71	0,29	0,59	0,67
Ag05	0,30	0,33	0,44	0,25	I	0,18	0,20	0,23	0,26	0,45	0,20	0,30	0,15	0,22	0,21
Ag06	0,19	0,37	0,19	0,45	0,18	I	0,40	0,50	0,40	0,09	0,64	0,52	0,27	0,35	0,33
Ag07	0,35	0,48	0,28	0,31	0,20	0,40	I	0,32	0,53	0,13	0,41	0,44	0,31	0,32	0,33
Ag08	0,20	0,42	0,29	0,55	0,23	0,50	0,32	I	0,37	0,11	0,67	0,60	0,36	0,49	0,40
Ag09	0,52	0,54	0,23	0,29	0,26	0,40	0,53	0,37	I	0,01	0,43	0,39	0,40	0,29	0,29
Ag10	0,17	0,29	0,67	0,31	0,45	0,09	0,13	0,11	0,01	I	0,08	0,37	0,10	0,29	0,31
Ag11	0,30	0,53	0,28	0,62	0,20	0,64	0,41	0,67	0,43	0,08	I	0,67	0,42	0,51	0,46
Ag12	0,34	0,59	0,58	0,71	0,30	0,52	0,44	0,60	0,39	0,37	0,67	I	0,39	0,59	0,62
Ag13	0,32	0,41	0,25	0,29	0,15	0,27	0,31	0,36	0,40	0,10	0,42	0,39	I	0,34	0,32
Ag14	0,19	0,43	0,38	0,59	0,22	0,35	0,32	0,49	0,29	0,29	0,51	0,59	0,34	I	0,51
Ag15	0,22	0,49	0,44	0,67	0,21	0,33	0,33	0,40	0,29	0,31	0,46	0,62	0,32	0,51	I

As agências com menores índices de correlação foram escolhidas como candidatas a serem modeladas. Assim, separou-se as dependências cujo índice de correlação é igual ou inferior a 0,10.

- Agências 06 e 10: correlação de 0,09;
- Agências 09 e 10: correlação de 0,01;
- Agências 10 e 11: correlação de 0,08;
- Agências 10 e 13: correlação de 0,10.

Posteriormente, analisando o gráfico de movimentação da agência 10, foi detectada uma drástica mudança de comportamento da série de dezembro de 1999 em diante, tendo um padrão diferente a partir daquele momento. Pesquisando as causas de tal alteração descobriu-se que nesta época ocorreu a terceirização dos caixas-eletrônicos da agência, passando a serem administrados por terceiros. Assim, os saques efetuados naquelas máquinas não eram mais contabilizados localmente, gerando uma mudança de comportamento na série.

Todas as agências do município do Rio sofreram o mesmo tratamento, ou seja, tiveram os seus caixas terceirizados, iniciando o processo em dezembro de 1999. Com isto a movimentação de numerário nestas dependências sofreu um forte impacto, alterando o seu comportamento. Como, no momento, foge ao escopo deste trabalho avaliar esta situação, não se trabalhou com nenhuma agência nesta situação.

Em virtude do fato acima, excluiu-se a agência 10 do grupo, ficando-se com as agências 06, 09, 11 e 13. Ainda havia a necessidade de se reduzir o conjunto pois o objetivo era de se modelar apenas 3 agências. A 13 foi a primeira escolhida, pois além de possuir baixa correlação com as demais, tem movimento “bem comportado”, ou seja, possui um dos menores desvios padrões de todo o conjunto e é localizada em cidade sem influência turística significativa, de porte pequeno e com média de movimentação de numerário baixa. Com todos estes atributos, espera-se tê-la como referência para a modelagem da rede.

Verificou-se também que a correlação entre a agência 06 e 11 era muito alta (0,64) e, por este motivo eliminou-se a dependência 11. Ficou-se, neste ponto com as agências 06, 09 e 13 para se trabalhar. Entretanto as filiais 09 e 13 são muito similares entre si, no tocante à média, valores mínimos, máximos e desvio-padrão. Assim, retirou-se a agência 09 do grupo e incluiu-se a agência 02, chegando-se ao conjunto

final. A substituição da agência 09 pela 02 ocorreu em função do fato de a ag. 02 estar localizada em região turística e ser, pelo seu porte, monitorada, constantemente, pelos analistas do Banco. Assim, o modelo criado pode ser avaliado em todos os seus aspectos. As correlações entre as agências selecionadas são:

- Agências 02 e 06: correlação de 0,37;
- Agências 02 e 13: correlação de 0,41; e
- Agências 06 e 13: correlação de 0,27.

2.3.1. O tratamento das agências 02, 06 e 13

A fim de melhor visualizar o comportamento das agências selecionadas, foram plotados os gráficos de todo o movimento de cada dependência.

Vê-se que a agência 02 possui como característica o fato de a maioria dos pontos estar próximo de 0 (média = -14,42) ou seja, o seu movimento é equilibrado. Alguns pontos estão fora do intervalo de [-400, 400] e muito afastados da média, demonstrando com isto alguma oscilação no movimento.

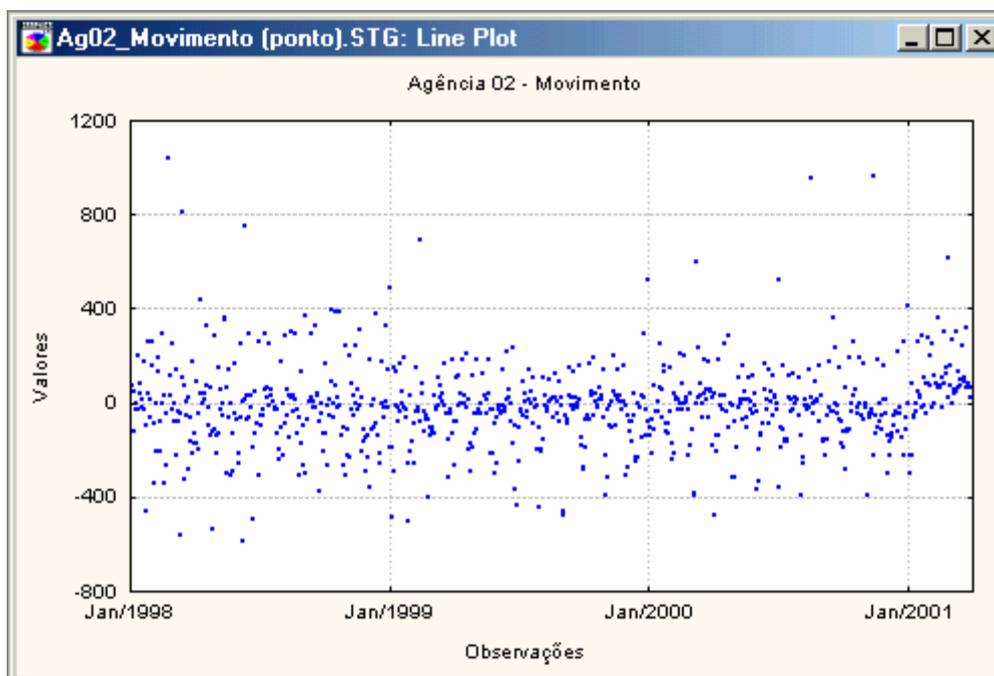


Gráfico 1: Agência 02 – representação de toda a série.

Já a agência 06 está caracterizada, através da observação do gráfico, como uma dependência recebedora, pois conforme se verifica a maioria dos seus movimentos

encontra-se acima de zero, demonstrando que os depósitos são superiores aos saques. Por sua vez os pontos estão muito dispersos, fazendo com que o desvio padrão seja alto.

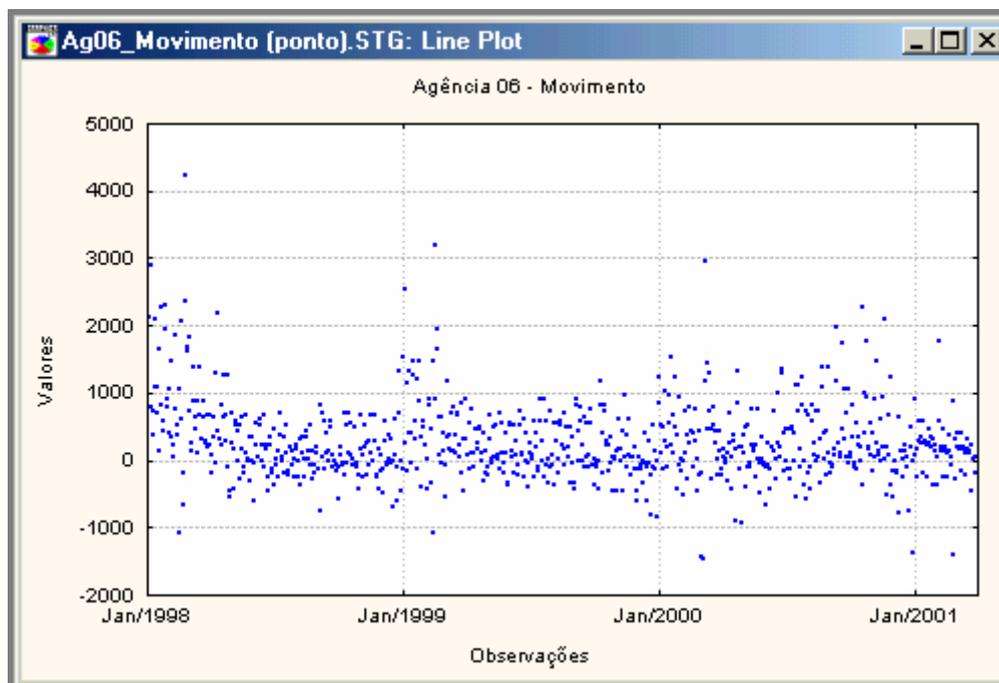


Gráfico 2: Agência 06 – representação de toda a série.

A agência 13 é claramente uma agência recebedora com a maioria do movimento de depósitos bem uniforme. Tirando um ponto espúrio verificado em 01/12/2000 os demais estão praticamente todos na faixa que vai de -100 até 250.

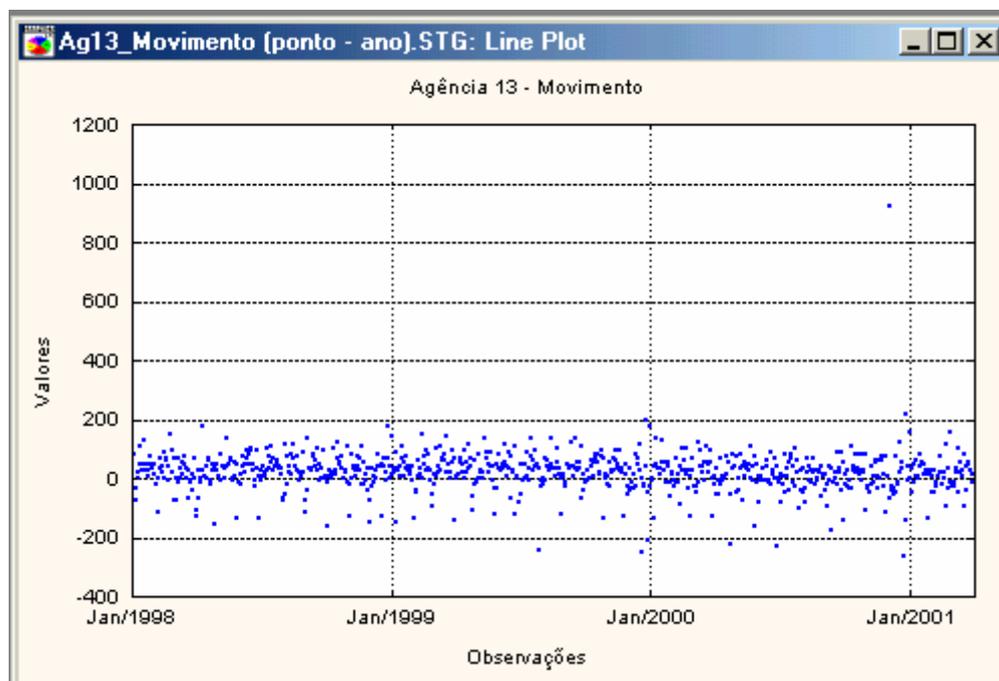


Gráfico 3: Agência 13 – representação de toda a série.

2.3.2. Análise Mensal dos Movimentos das Agências 02, 06 e 13

A seguir foram construídos gráficos com enfoque mensal. Neste caso usou-se o eixo das abscissas – eixo x – para representar os meses e o eixo das ordenadas os valores. Para cada mês no eixo das abscissas marcou-se os valores relativos a todos os dias do mês, ou seja, para o ponto 1 do eixo x, representando o mês de janeiro de 1998, têm-se, na vertical, os valores de todos os dias deste mês, e assim sucessivamente. Representou-se, também, em forma de losango, as médias mensais e, por fim, com linhas horizontais os valores referentes a 1 desvio e 2 desvios padrões da série.

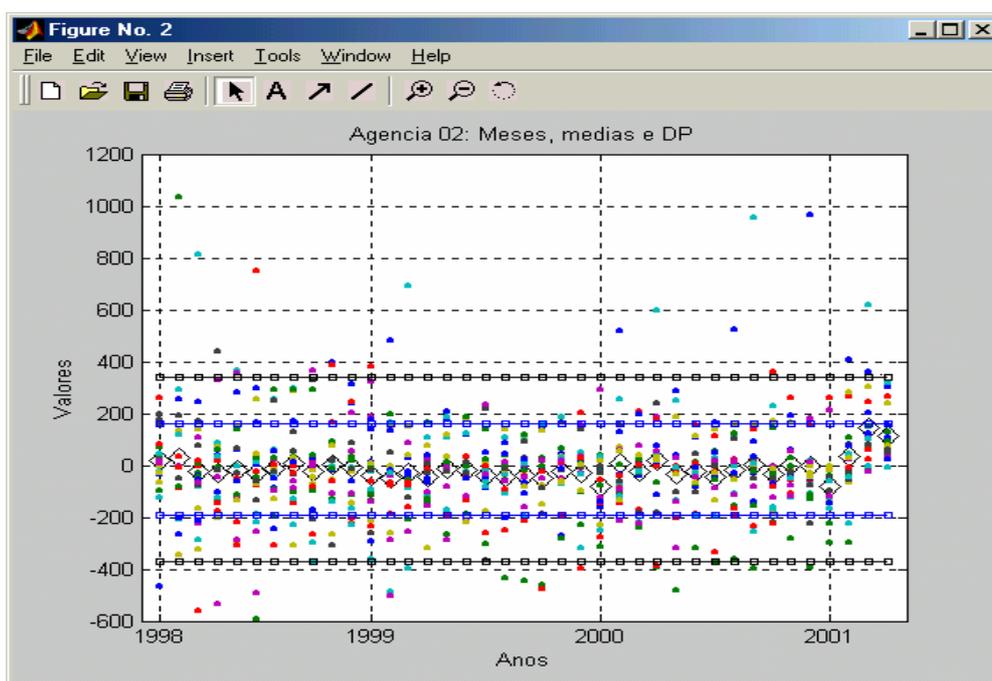


Gráfico 4: Agência 02 - Valores mensais, médias e desvio padrão. O losango representa a média aritmética dos pontos no eixo das ordenadas (y) e as linhas horizontais correspondem aos valores referentes a 1 desvio e 2 desvios padrões da série.

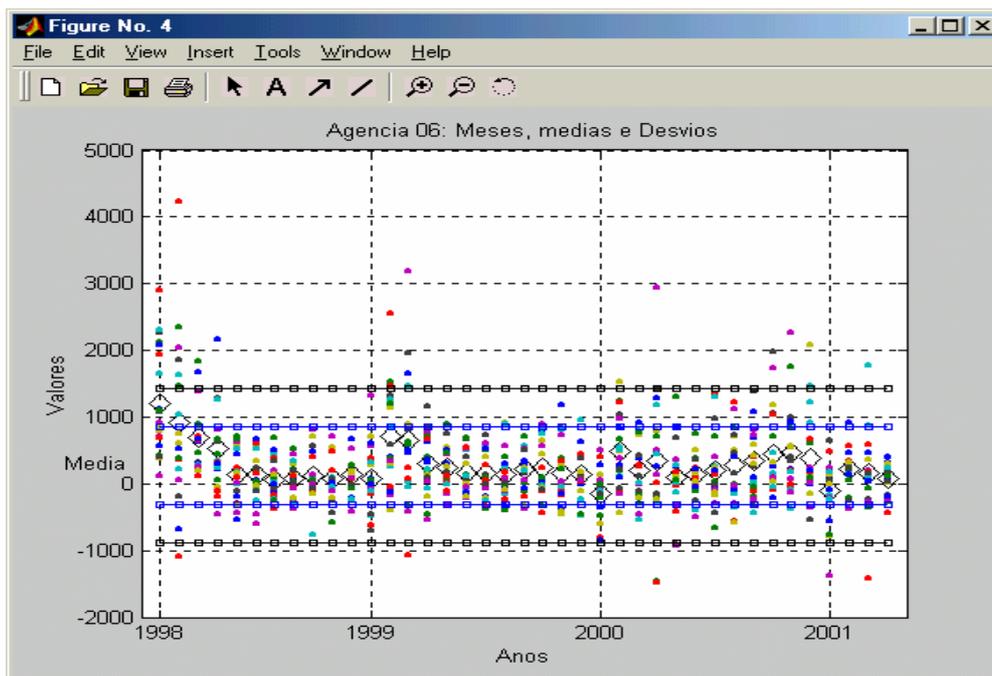


Gráfico 5: Agência 06 - Valores mensais, médias e desvio padrão. O losango representa a média aritmética dos pontos no eixo das ordenadas (y) e as linhas horizontais correspondem aos valores referentes a 1 desvio e 2 desvios padrões da série.

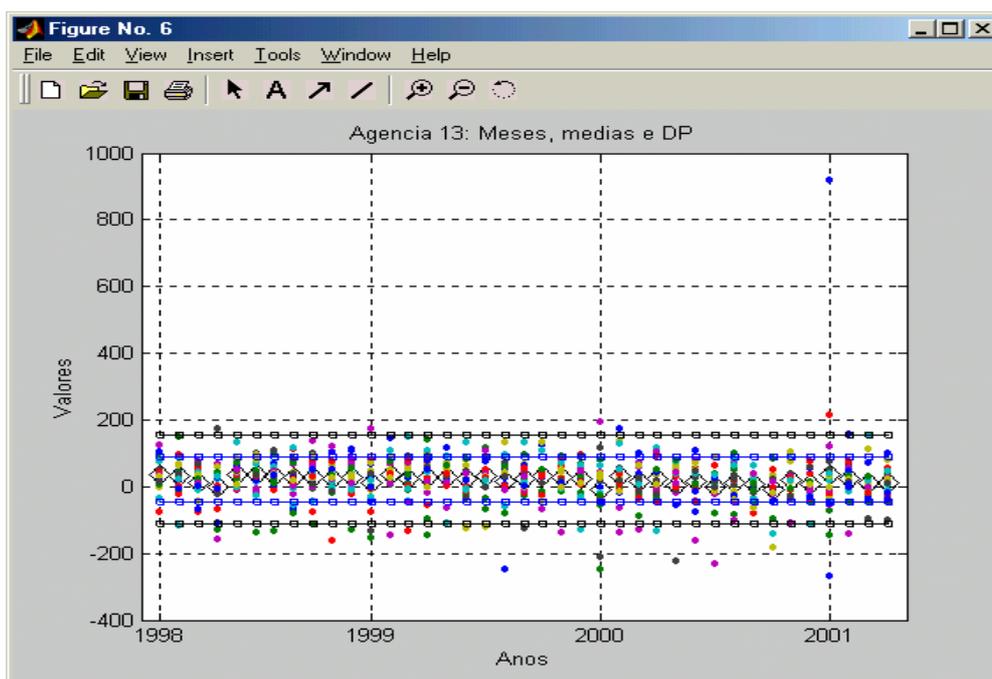


Gráfico 6: Agência 13 - Valores mensais, médias e desvio padrão. O losango representa a média aritmética dos pontos no eixo das ordenadas (y) e as linhas horizontais correspondem aos valores referentes a 1 desvio e 2 desvios padrões da série.

Para uma distribuição normal é esperado que 95,5% dos pontos estejam contidos na faixa de $[\bar{x} - 2s, \bar{x} + 2s]$, onde \bar{x} corresponde à média e s ao desvio padrão (STEVENSON, 1986, p. 140). Verifica-se que tanto a agência 02 como a 06 têm 769

pontos dentro do intervalo (42 pontos fora), representando 94,8% dos pontos e a agência 13 possui 771 pontos no intervalo (40 pontos fora), que corresponde a 95,1% dos registros. A mesma análise foi feita para a situação de 1 desvio padrão, onde, segundo a mesma referência (STEVENSON, 1986, p. 140), 68,0% dos dados devem estar na faixa de $[\bar{x} - s, \bar{x} + s]$, se houver uma distribuição normal. Viu-se, neste caso, que a agência 02 tem 612 pontos no intervalo citado (199 fora), representando 75,5%, já a agência 06 possui 647 pontos no intervalo (164 fora) o que significa que 79,8% estão inclusos no intervalo e a agência 13 possui 662 pontos pertencentes ao intervalo de 1 desvio padrão (149 fora), indicando que 81,6% dos pontos pertencem ao conjunto. Posteriormente, se fará análise mais detalhada dos registros que saíram do padrão de distribuição normal dos dados, entretanto pode-se adiantar que os dias onde se verificaram registros fora do intervalo de 2DP não são coincidentes para todas as agências pesquisadas – agências 02, 06 e 13 –, ou seja, a ocorrência de *outlier* em uma agência não corresponde, necessariamente, a de outra.

A seguir são apresentados gráficos do desvio padrão mensal das séries e o intervalo compreendido pela soma e subtração do valor médio mensal com um desvio padrão.

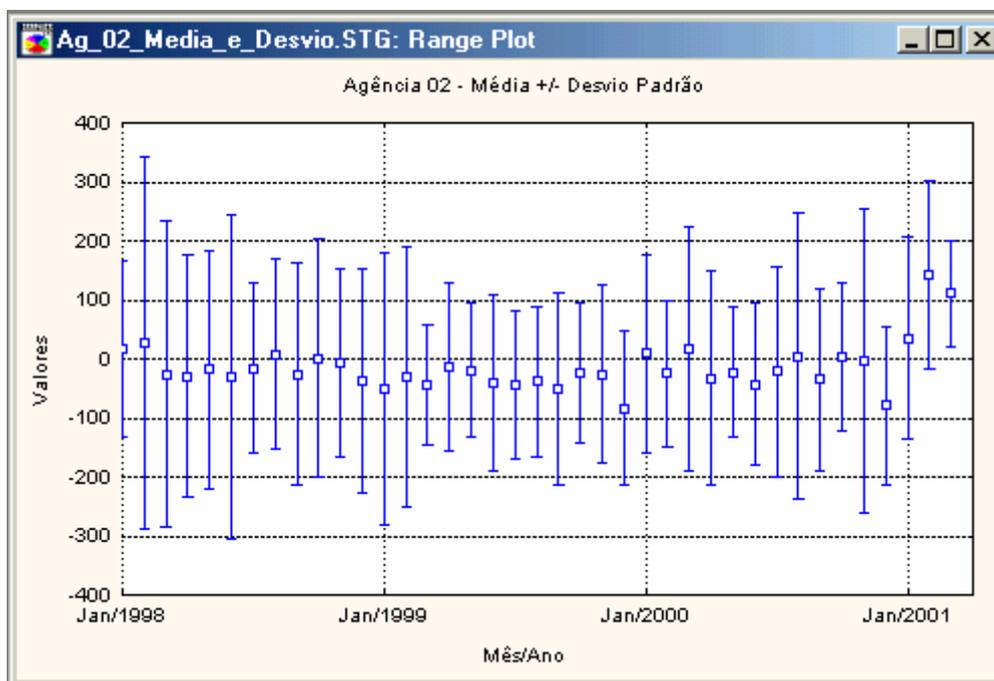


Gráfico 7: Agência 02 – Média +/- desvio padrão.

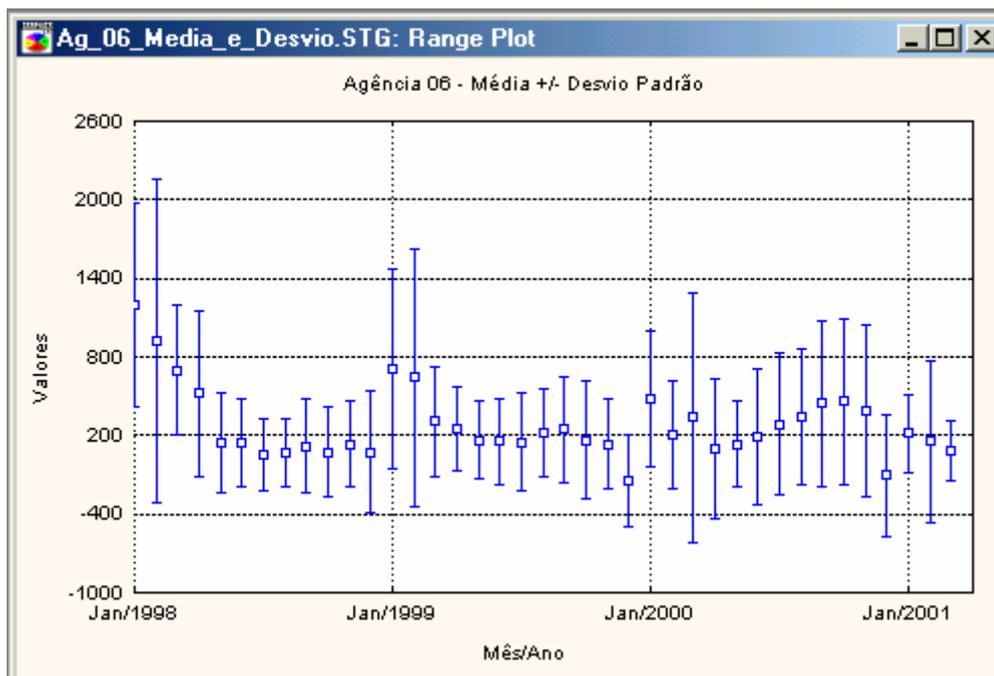


Gráfico 8: Agência 06 – Média +/- desvio padrão.

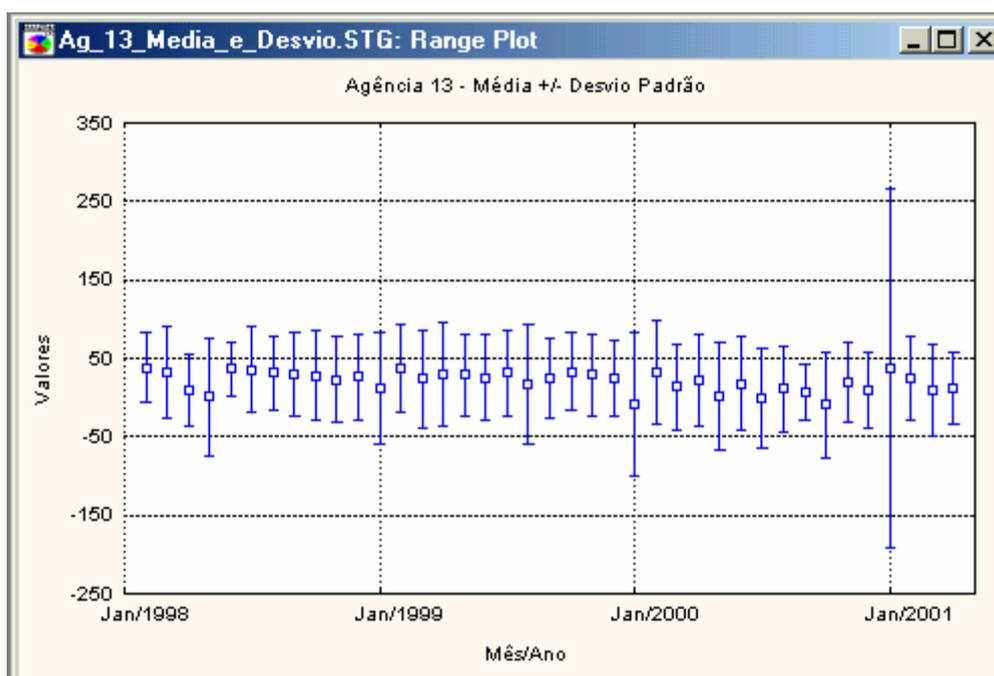


Gráfico 9: Agência 13 – Média +/- desvio padrão.

A agência 02, seguida pela agência 06 são as que mostraram as maiores dispersões dos dados. Tratam-se de agências localizadas em balneários onde ocorre um aumento do movimento no verão, ocasionando grandes oscilações na movimentação do numerário. Pode-se verificar que os períodos onde ocorrem as variações não são

idênticos, ou seja, enquanto a agência 02 apresenta os maiores desvios no ano de 1999, e alguns meses de 2000; na agência 06 as ocorrências são no início de 1998, início de 1999 e o segundo semestre de 2000; já a agência 13, por sua vez, é a mais bem comportada das 3 e seus valores quase não apresentam grandes oscilações. O desvio padrão desta agência é o mais baixo de todas.

2.3.3. Análise Semanal dos Movimentos das Agências 02, 06 e 13

Busca-se, agora, verificar os movimentos das agências sob o foco dos dias da semana, ou seja, nesta análise a preocupação está voltada para estudar o comportamento existente durante a semana. A expectativa, neste caso, é de se verificar fortes recebimentos nos primeiros dias da semana (segunda-feira e terça-feira), equilíbrio na quarta-feira e fortes pagamentos nos últimos dias da semana (quinta-feira e sexta-feira) pois acredita-se que hajam depósitos de comércio no início da semana e saques da população no últimos dias da semana.

A primeira agência a ser tratada é a 02 e o quadro com os valores mínimos, máximos, médias e desvios padrões por dia da semana estão em seguida.

Quadro 8: Agência 02. Valores mínimos, máximos, médias e desvios padrões dos dias da semana.

	Agência 02			
	Valor Mínimo	Valor Máximo	Média	Desvio Padrão
Segunda-Feira	-333,99	966,67	164,54	171,57
Terça-Feira	-479,67	956,67	-19,66	150,82
Quarta-Feira	-395,36	1035,59	-31,40	163,08
Quinta-Feira	-535,00	201,70	-69,73	125,17
Sexta-Feira	-590,20	360,07	-110,09	141,22

Foi plotado, também, o gráfico com os valores semanais. Para tanto, adotou-se os mesmos procedimentos utilizados anteriormente, ou seja, no eixo das abscissas – eixo x – foram colocados todos os pontos relativos ao dia da semana a que se refere. Assim, no ponto 1 têm-se, na vertical, todos os valores relativos às segundas-feiras, no ponto 2 os valores das terças-feiras e assim sucessivamente. O eixo das coordenadas – eixo y – apresenta os valores das movimentações verificadas no dia.

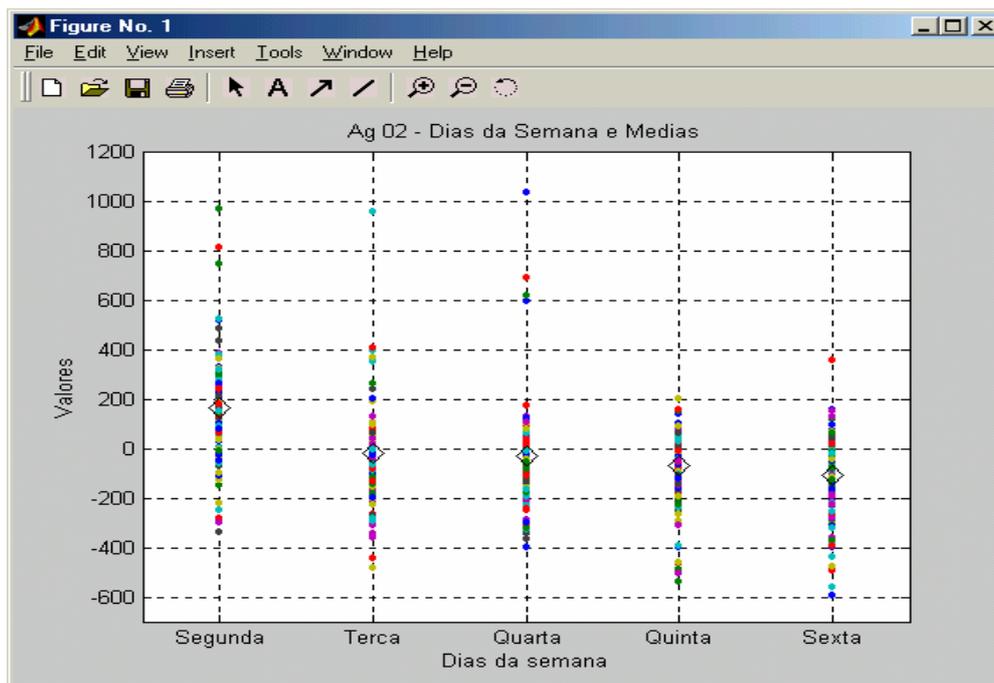


Gráfico 10: Agência 02. Dias da semana e médias.

Pode-se afirmar que a agência 02 não se comporta conforme a expectativa, pois nas terças, quartas e quintas o seu movimento é bem equilibrado, tendo médias baixas. Isto contraria a afirmação de que nas terças-feiras a agência ainda sofre influência das segundas-feiras e tem movimento de depósitos. Este movimento não deveria ser tão alto como os das segundas-feiras mas, ainda assim, recebedor. Não foi o verificado, pois nas terças-feiras a média da agência é pagadora.

As quartas-feiras que deveriam ter um movimento bem equilibrado, ou seja, com médias próximas de zero, apresentam-se como pagadoras. As quintas e sextas-feiras seguem o padrão esperado que são os dias onde os saques excedem os depósitos e com as sextas-feiras tendo movimentos mais fortes que as quintas-feiras.

A agência 06 só apresentou comportamento coerente com o esperado nas segundas-feiras e sextas-feiras, onde se teve no primeiro um forte movimento de depósitos e nas sextas-feiras de saques. Cabe destacar que nas terças-feiras os movimentos de depósitos foram superiores aos das segundas, fato este totalmente anômalo. Outra característica que chamou a atenção diz respeito ao movimento de depósitos verificado nas quintas-feiras, dia tipicamente de saques.

Quadro 9: Agência 06. Valores mínimos, máximos, médias e desvios padrões dos dias da semana.

	Agência 06			
	Valor Mínimo	Valor Máximo	Média	Desvio Padrão
Segunda-Feira	-532,59	2126,00	543,38	412,11
Terça-Feira	-369,01	2905,64	659,62	569,20
Quarta-Feira	-619,56	2545,79	254,24	490,32
Quinta-Feira	-1448,85	4227,61	102,92	558,31
Sexta-Feira	-1470,37	1645,40	-179,49	405,68

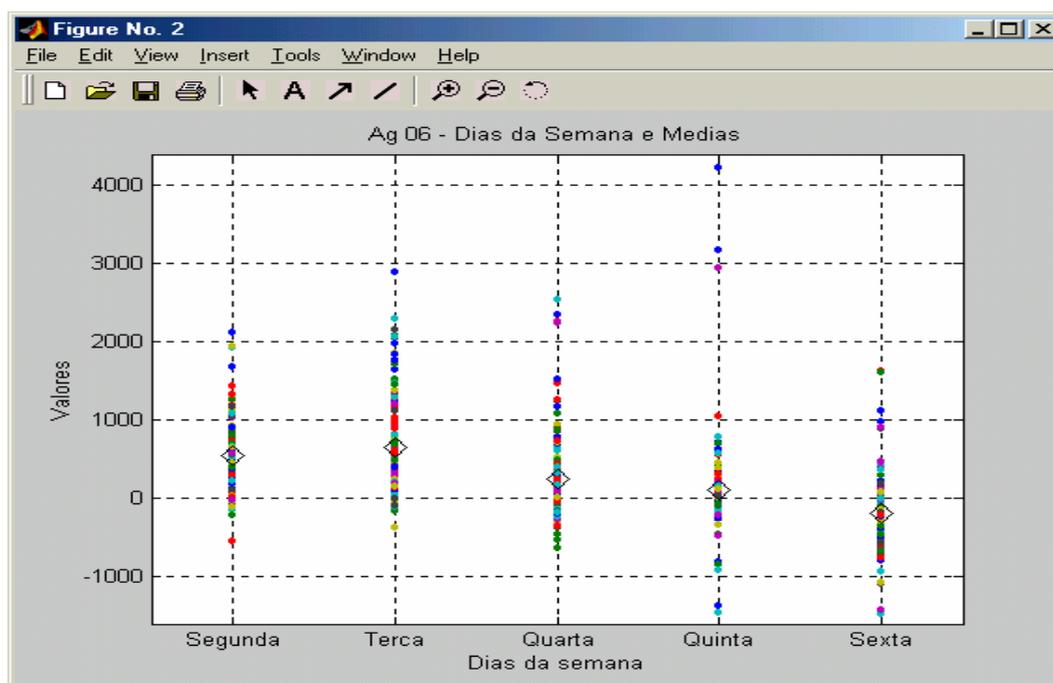


Gráfico 11: Agência 06. Dias da semana e médias.

Por fim, a agência 13 que anteriormente era considerada como a dependência mais dentro do perfil esperado, se apresentou com características distintas das desejadas, quais sejam: as quintas-feiras, contrariamente ao esperado, apresenta depósitos, ao invés de saques e nas sextas-feiras os saques são, em média, fracos, com características de quartas-feiras, onde se tem movimento equilibrado.

Quadro 10: Agência 13. Valores mínimos, máximos, médias e desvios padrões dos dias da semana.

	Agência 13			
	Valor Mínimo	Valor Máximo	Média	Desvio Padrão
Segunda-Feira	-178,96	195,71	66,66	55,77
Terça-Feira	-135,07	214,11	23,82	51,55
Quarta-Feira	-248,10	154,61	8,40	57,88
Quinta-Feira	-265,84	95,71	6,05	51,47
Sexta-Feira	-247,34	919,40	-0,10	88,11

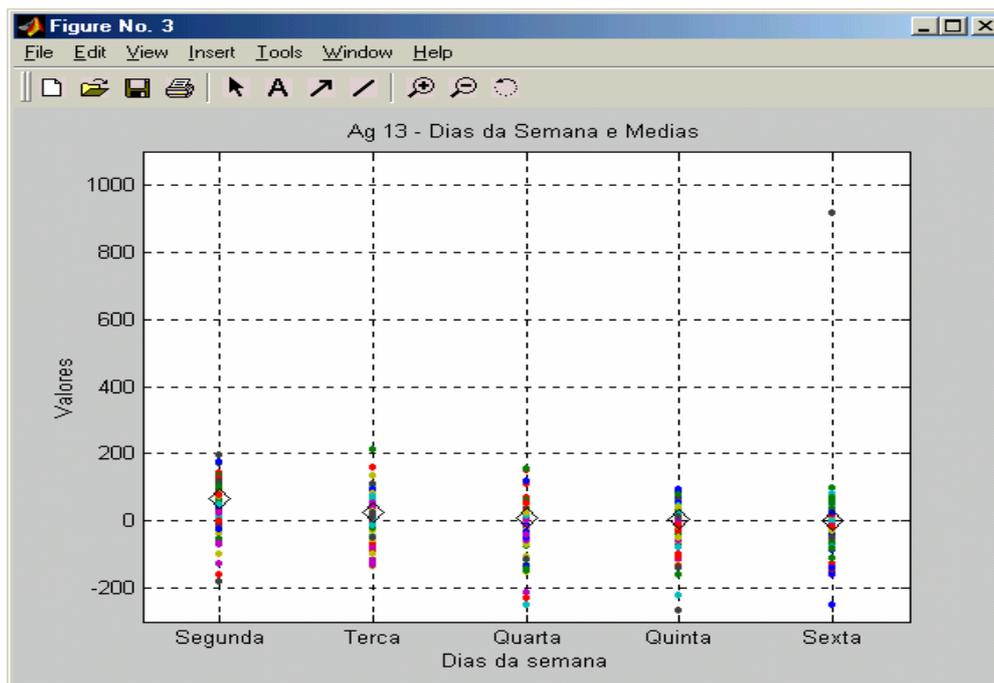


Gráfico 12: Agência 13. Dias da semana e médias.

Conclui-se, após o estudo acima, que somente as segundas-feiras e sextas-feiras mantêm um padrão de comportamento do tipo depósitos nas segundas-feiras e saques nas sextas-feiras. Ainda assim, conforme se observou na agência 06, os depósitos das segundas-feiras não necessariamente são os maiores da semana e os saques verificados nas sextas-feiras da agência 13 não são tão vultosos.

Tais fatos, mencionados no parágrafo anterior, levam a acreditar que esses padrões de comportamento devem ser vistos com cautela, pois eles não seguem uma regra fixa. Acredita-se que muita das distorções verificada nas análises se deve ao fato dos feriados existentes, pois, nestes casos, o movimento dos dias que antecedem e sucedem o feriado são totalmente atípicos. Feriados nas terças-feiras e quintas-feiras fazem com que os movimentos das quartas-feiras sejam totalmente divergentes daquele característico deste dia. Além disso, as segundas-feiras e sextas-feiras que ficam entre os feriados também mudam de comportamento, pois grande parte da população tem por hábito “enforçar” este dia, diminuindo, conseqüentemente, o movimento bancário.

O estudo da influência dos feriados no comportamento da série assim como a geração de *outliers* decorrentes deste fato estão abrangidos no capítulo 4 – Análise das Séries em Estudo.

2.3.4. Análise Diária dos Movimentos das Agências 02, 06 e 13

Acredita-se que o comportamento típico de uma agência é pautado em fortes pagamentos no início de cada mês (5 primeiros dias úteis), crescendo suavemente para o equilíbrio nos demais dias. Baseado nesta expectativa gerou-se gráficos por dias, ou seja, o eixo das abscissas marca, a cada ponto, um determinado dia útil (1º. dia útil, 2º. dia útil etc) e o eixo das ordenadas os valores movimentados. Foi criado, também, um losango para registrar o valor médio de cada dia útil do mês.

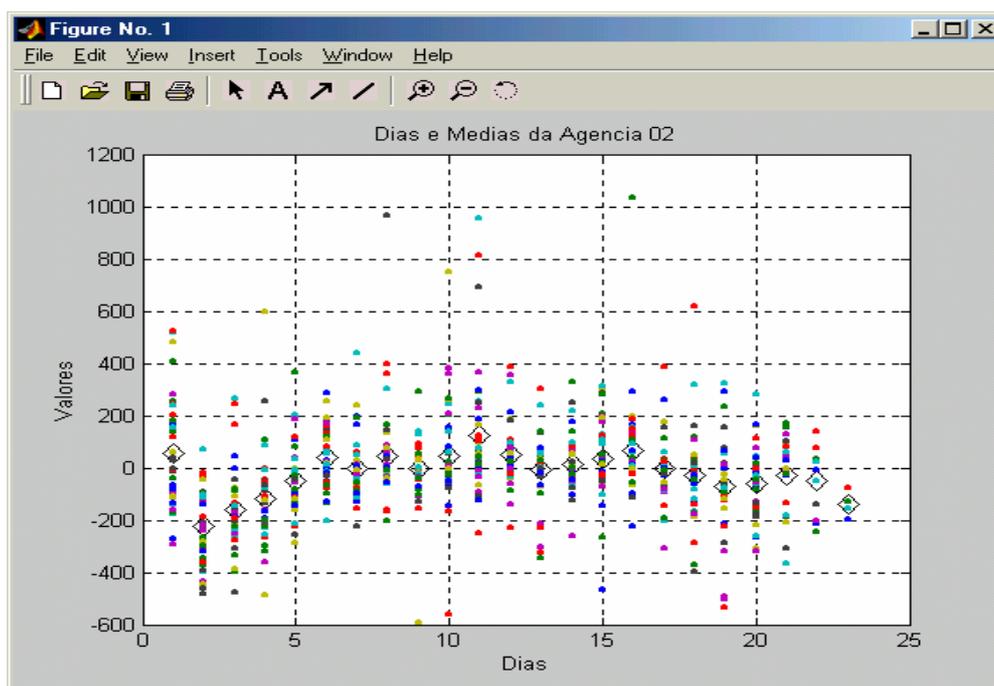


Gráfico 13: Agência 02. Dias úteis do mês e médias.

Vê-se no gráfico da agência 02 que, na média, efetivamente há um forte pagamento no início do mês, a exceção do primeiro dia útil onde a média foi positiva. A partir do 6º. dia útil o movimento tende a se estabilizar, sofrendo uma leve queda por volta do 20º. dia útil do mês. A seguir tem-se quadro com os valores mínimos, máximos, médias e desvios padrões de cada dia útil do mês.

Quadro 11: Agência 02. Valores mínimos, máximos, médias e desvios padrões de cada dia útil do mês.

	Agência 02			
	Valor Mínimo	Valor Máximo	Média	Desvio Padrão
Dia 01	-292,47	522,84	55,55	209,87
Dia 02	-479,67	71,83	-225,39	145,58
Dia 03	-477,70	265,86	-160,85	155,97
Dia 04	-486,21	598,65	-118,37	177,16
Dia 05	-287,41	366,92	-52,40	131,01
Dia 06	-200,39	286,71	38,90	116,25
Dia 07	-221,63	438,21	-0,64	119,17
Dia 08	-202,71	966,67	45,81	194,76
Dia 09	-590,20	293,83	-3,73	132,76
Dia 10	-558,56	748,43	43,07	198,73
Dia 11	-249,69	956,67	124,95	241,43
Dia 12	-228,36	384,68	49,32	119,69
Dia 13	-343,86	304,51	-8,52	132,64
Dia 14	-259,82	326,36	11,34	105,99
Dia 15	-464,82	312,03	34,39	151,12
Dia 16	-225,57	1035,59	66,61	186,31
Dia 17	-307,14	385,71	-1,19	128,17
Dia 18	-399,27	618,99	-31,54	169,82
Dia 19	-535,00	325,63	-73,46	195,77
Dia 20	-319,16	279,94	-62,16	139,33
Dia 21	-364,06	173,02	-28,20	144,53
Dia 22	-244,56	140,00	-50,94	129,01
Dia 23	-196,13	-75,72	-138,96	50,70

A próxima agência a ser tratada é a 06 e ela se apresenta, na média, recebedora durante praticamente todos os dias do mês. Nem mesmo nos 5 primeiros dias pode-se dizer que se comporta segundo os padrões que acreditava-se ser inerente a todas as dependências, ou seja, pagadora. Verifica-se que os desvios padrões dos dias úteis desta agência são extremamente altos, caracterizando grandes oscilações. Outro fato passível de comentário é a grande diferença existente entre os valores mínimos e máximos dos dias. Ora, se a agência devia se comportar como pagadora nos primeiros dias do mês não justifica termos valores máximos tão altos. Vê-se, inclusive, que dos cinco maiores valores máximos dos dias (1º. dia útil, 3º. dia útil, 5º. dia útil, 12º. dia útil e 17º. dia útil), três deles (1º. dia útil, 3º. dia útil e 5º. dia útil) fazem parte do grupo de dias que deveriam ser pagadores.

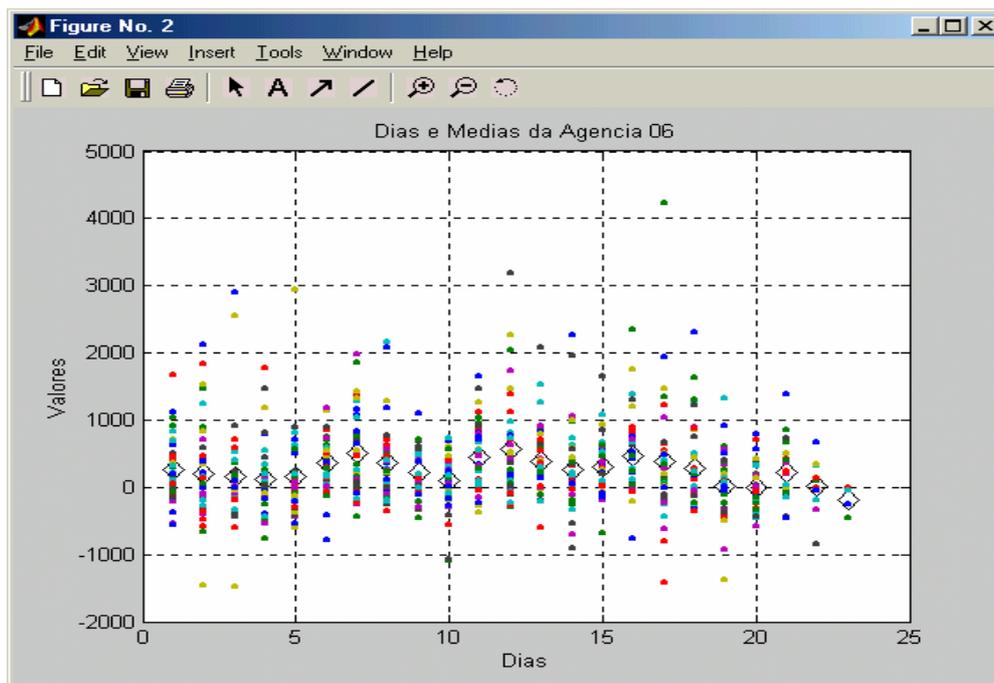


Gráfico 14: Agência 06. Dias úteis do mês e médias.

Quadro 12: Agência 06. Valores mínimos, máximos, médias e desvios padrões de cada dia útil do mês.

	Agência 06			
	Valor Mínimo	Valor Máximo	Média	Desvio Padrão
Dia 01	-550,00	1681,45	268,53	478,31
Dia 02	-1448,85	2126,00	197,26	713,73
Dia 03	-1470,37	2905,64	153,83	718,67
Dia 04	-759,85	1776,16	125,39	523,67
Dia 05	-593,35	2952,06	158,16	609,21
Dia 06	-779,07	1187,86	355,38	395,67
Dia 07	-425,36	1980,57	499,16	602,99
Dia 08	-347,61	2170,93	365,37	543,18
Dia 09	-449,31	1098,88	214,84	331,69
Dia 10	-1090,00	741,86	93,72	400,42
Dia 11	-369,01	1645,40	453,16	483,48
Dia 12	-283,39	3183,85	563,89	754,20
Dia 13	-595,41	2080,20	383,08	495,34
Dia 14	-905,65	2259,71	268,97	610,13
Dia 15	-670,57	1654,47	302,12	418,20
Dia 16	-757,50	2355,93	475,59	584,31
Dia 17	-1419,82	4227,61	376,26	897,05
Dia 18	-359,57	2300,17	289,98	600,28
Dia 19	-1370,65	1322,76	14,05	496,47
Dia 20	-570,12	794,53	3,77	333,53
Dia 21	-446,88	1384,99	215,33	414,83
Dia 22	-844,92	666,00	12,89	394,92
Dia 23	-453,53	0,00	-188,36	209,05

Por último, são apresentados os dados estatísticos diários e o gráfico da agência 13. Conforme comentado anteriormente, é uma dependência equilibrada sendo

que os desvios padrões são baixos e as médias de todos os dias do mês próximas de zero. Este último item também está em desacordo com a expectativa, pois para que a agência fosse fortemente pagadora, as médias dos primeiros dias úteis deveriam ser altas e negativas, o que não ocorre.

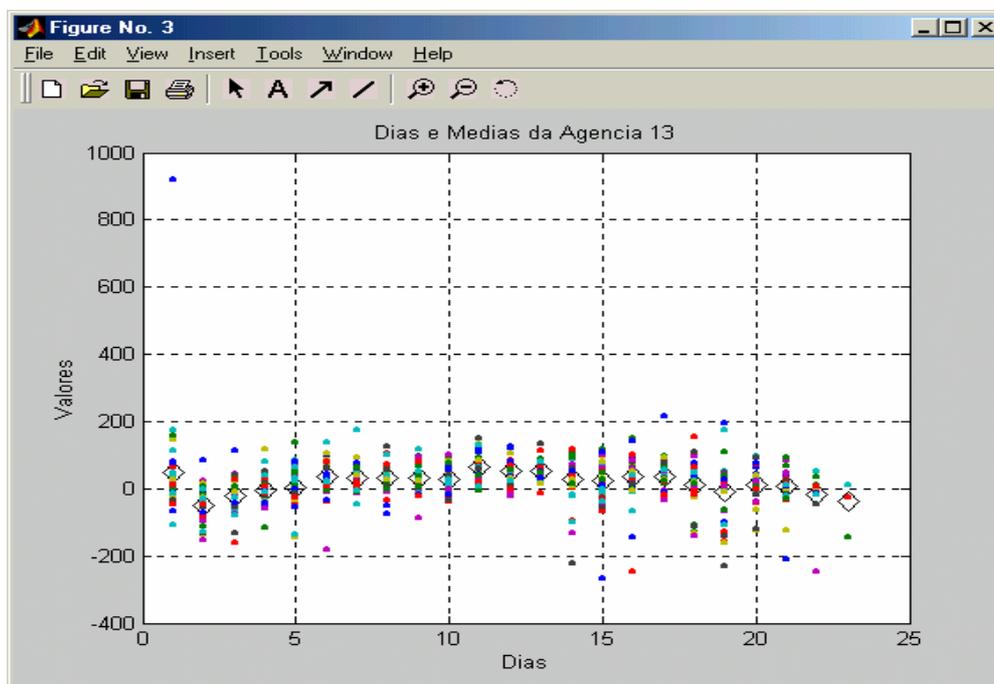


Gráfico 15: Agência 13. Dias úteis do mês e médias.

Quadro 13: Agência 13. Valores mínimos, máximos, médias e desvios padrões de cada dia útil do mês

	Agência 13			
	Valor Mínimo	Valor Máximo	Média	Desvio Padrão
Dia 01	-105,43	919,40	48,74	154,52
Dia 02	-152,39	84,06	-49,32	49,88
Dia 03	-159,67	112,85	-22,39	49,88
Dia 04	-115,33	119,19	-3,38	39,72
Dia 05	-145,69	136,61	5,23	53,30
Dia 06	-178,96	137,47	34,03	53,02
Dia 07	-46,49	173,53	31,59	35,67
Dia 08	-74,36	124,34	32,46	39,09
Dia 09	-85,75	116,46	29,85	37,81
Dia 10	-37,40	101,88	29,26	37,76
Dia 11	-3,79	150,80	64,42	41,18
Dia 12	-22,41	126,35	50,41	35,97
Dia 13	-14,44	133,76	52,13	29,73
Dia 14	-221,20	116,00	25,98	67,13
Dia 15	-265,84	117,96	22,17	67,17
Dia 16	-248,10	152,10	35,07	69,44
Dia 17	-33,11	214,11	37,22	43,21
Dia 18	-140,57	154,61	11,25	63,95
Dia 19	-228,56	195,71	-10,38	90,41
Dia 20	-125,06	95,93	10,94	51,84
Dia 21	-210,15	95,05	6,02	62,05
Dia 22	-247,34	51,44	-19,13	79,97
Dia 23	-145,61	10,59	-37,81	73,99

2.4. CONCLUSÕES

Esta dissertação trata tão-somente da previsão de numerário, ou seja, a resultante dos depósitos e saques realizados pelo público em dinheiro. Cheques e outros documentos financeiros não fazem parte do escopo do trabalho.

É senso comum imaginar que a movimentação de numerário em agências bancárias segue o padrão de depósitos no início da semana, caminhando-se para o equilíbrio no meio da semana e saques nos últimos dois dias da semana. O primeiro dia da semana é onde se encontram os maiores depósitos, o segundo ainda sofre influência do início da semana e, portanto, ainda se caracteriza por movimento recebedor; a quarta-feira é vista como sendo equilibrada com resultante entre os saques e depósitos do público próximo de zero. A partir da quinta-feira a curva representativa do movimento passa a ser negativa e crescente, sendo os maiores saques verificados na sexta-feira.

Das agências escolhidas para esta dissertação – 02, 06 e 13 –, nenhuma manteve o padrão típico esperado, quais sejam: fortes depósitos nas segundas-feiras, terças-feiras com depósitos menores, quartas-feiras equilibradas, saques nas quintas-feiras e fortes saques nas sextas-feiras.

Sob a ótica do mês os primeiros 5 dias úteis são considerados de movimento pagador e os demais com tendências ao equilíbrio. Cabe comentar que não há contradição entre as afirmativas anteriores pois o fato de se ter movimentos pagadores não invalida que, nas segundas-feiras, mesmo estando inserida nos primeiros 5 dias úteis, haja, em alguns casos, movimento recebedor na agência.

Também sob o aspecto do mês o comportamento das agências não se mostrou totalmente dentro dos padrões definidos. A agência 02 é, na média, recebedora no 1º. dia útil do mês, sendo pagadora no 2º., 3º., 4º. e 5º. dias úteis. O resto do mês ela se comporta em equilíbrio. A agência 06 apresenta-se como recebedora em praticamente todo o mês e a agência 13 com leve tendência pagadora no início e fim do mês, tendo característica equilibrada nos demais dias.

Algumas das distorções observadas nas análises se devem aos feriados existentes. Feriados nas terças-feiras e nas quintas-feiras são comumente emendados com os dias que antecedem/sucedem o feriado gerando, em consequência, movimentos atípicos nos dias posteriores aos feriados. Os anos de 1999 e 2000 tiveram, cada um,

quatro feriados com esta característica. Outro período importante que gera distorção é o carnaval que influencia praticamente toda a semana anterior e os três dias seguintes ao evento (quarta-feira de cinzas, quinta-feira e sexta-feira após o carnaval). No capítulo 4 será abordada a questão relativa aos feriados e qual o tratamento adotado.

As 15 agências constantes do conjunto fornecido pelo Banco para estudo foram avaliadas. Entretanto, baseado nas análises promovidas alguns conceitos amplamente aceitos não foram confirmados, tais como:

- Contrariamente ao que se espera não se verificou a característica pagadora nos primeiros cinco dias úteis de cada mês, em todas as agências. Algumas delas, inclusive, foram fortemente recebedoras neste período, a exemplo da agência 06;
- Não se observou relação entre agências situadas em localidades com o mesmo perfil, a exemplo de agências em municípios agrícolas, turísticos ou outro;
- Agências similares, localizadas no mesmo município também não apresentaram correlação, como foi o caso de Copacabana-Rio de Janeiro e Barra da Tijuca – Rio de Janeiro;
- O fato das segundas-feiras serem dias basicamente recebedores e as sextas-feiras pagadores também não correspondeu à realidade. Houve várias situações em que está lógica não foi verificada;
- A ocorrência de *outliers* não seguiu nenhum padrão. Não foram encontradas nenhuma situação em que todas as agências, no mesmo dia, tivessem valores elevados. Cabe frisar que foram considerados *outliers* os valores superiores a 2 desvios padrões. Entretanto, estes dados não podem ser considerados como erro pois foram conferidos e correspondem, efetivamente, ao movimento verificado no dia da ocorrência.

2.5. HIPÓTESES

Verificou-se, em todas as séries estudadas, um grande número de oscilações no comportamento dos dados, ou seja, tem-se em um dia “d” valores positivos e, logo em seguida, em “d + 1”, valores negativos, ou vice-versa. Este fenômeno evidencia a

existência e relevância de componentes de alta frequência, dificultando o mapeamento da série. Além da alta frequência verificada também observou-se a existência de significativa dispersão dos dados bem como de valores extremamente elevados, tanto de saques como de depósitos, o que foi tratado como *outliers*.

Com base nas análises iniciais é de se supor que se tratam de séries não lineares onde a aplicação de modelos estatísticos convencionais não deve apresentar bons resultados. O uso, portanto, de técnicas de inteligência computacional deve ser investigado com vistas a obtenção de melhores resultados. Entretanto, mesmo neste caso, há a possibilidade de não se conseguir um bom desempenho em razão do excessivo número de *turnpoints*, de *outliers* e da grande dispersão dos valores.

Apesar não poder se afirmar, inequivocamente, a superioridade das redes neurais em comparação com os métodos estatísticos, os trabalhos de predição de comportamento futuro utilizando-se dessa técnica vêm obtendo melhores resultados que os métodos estatísticos (FERNANDES; NAVAU; PORTUGAL, 1996, p. 253) (BACHA; MEYER, 1992, p. II-442). Esta é, portanto, a principal motivação para se adotar as técnicas de inteligência computacional na série.

3. FERRAMENTAS DE ANÁLISE DE SÉRIES TEMPORAIS

Neste capítulo comentam-se as técnicas de pré-processamento de dados utilizadas na dissertação. No 2º item discorre-se sobre os conceitos estatísticos para, em seguida, apresentar algumas técnicas de transformação de dados. Nos itens 4 e 5 abordam-se os temas relativos ao tratamento de valores inexistentes e outliers. Por fim, a seção 6 encerra com a análise do capítulo.

3.1. INTRODUÇÃO

Um dos principais instrumentos extensamente utilizado na estatística é o modelo. Os modelos são versões simplificadas de algum problema ou situação da vida real. São usados para ilustrar certos aspectos da situação, evitando grande número de detalhes que talvez sejam irrelevantes para o problema (STEVENSON, 1986, p. 4). Portanto, a construção de um modelo privilegiará as informações mais relevantes do problema em detrimento daquelas que, num primeiro momento, se mostrem menos relevantes.

Ferramentas de análise são figuras estatísticas que visam representar o mundo real de forma simplificada, facilitando a análise e a compreensão do problema. Sob esta ótica encontram-se 3 grandes áreas onde se inserem as técnicas da estatística clássica:

- Estatística descritiva;
- Probabilidade; e
- Inferência estatística.

3.2. FERRAMENTAS DE ANÁLISE

Dependendo da quantidade de dados a serem analisados, pelo menos duas formas de tratamento podem ser realizadas: medidas de tendência central e agrupamento de dados. A primeira é freqüentemente usada para pequenos volumes e o agrupamento para grandes conjuntos. Estas técnicas não são excludentes, pode-se usá-las conjuntamente em qualquer massa de dados.

As medidas de tendência central como média, moda e mediana apresentam como característica o fato de servirem para tipificar um conjunto de dados, ou seja, apresentarem o valor mais típico do conjunto.

Já quando se trabalha com grandes massas de dados, deseja-se conhecer, principalmente, o comportamento do conjunto. Para isto, utilizam-se as técnicas de distribuições de frequência, intervalos etc.

3.2.1. Medidas de Tendência Central

Média Aritmética

A média, assim como a mediana e a moda são consideradas medidas de tendência central. Tais medidas têm como objetivo representar toda a massa de dados a partir de um único número.

A média aritmética é a medida de tendência central mais utilizada e é representada como:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{Eq. (2)}$$

sendo n o número total de observações e x_i o valor de cada observação.

A média possui particularidades que a torna útil para diversos propósitos (STEVENSON, 1986, p. 20):

- sempre haverá condições de ser calculada;
- seu valor é único em um determinado grupo;
- seu valor é modificado sempre que algum elemento novo for incluído;
- qualquer operação aritmética aplicada a cada valor do conjunto faz com que a média sofra esta mesma operação, ou seja, o valor da média ficará multiplicado por uma constante x se cada valor individualmente for multiplicado por esta mesma constante;
- a soma dos desvios dos números a partir da média é sempre zero.

Como desvantagem está o fato da média ser uma medida muito sensível aos limites extremos.

Média Móvel

Média móvel (MM) é a média aritmética das últimas k observações. Ao introduzir um novo valor na média, retira-se o mais antigo de tal forma que k seja sempre constante. Assim, vai-se caminhando através da coleção de dados.

Matematicamente, pode-se representar a média móvel como:

$$MM = \frac{1}{k} \sum_{i=t-k}^t Y_i \quad \text{Eq. (3)}$$

onde Y_i corresponde aos valores do conjunto, k a quantidade de observações que serão computadas e t o número de períodos de tempo.

Esta técnica é muito utilizada para regularização de dados e predição, sendo que uma das dificuldades na sua utilização é a escolha do tamanho de k , também conhecido como “janela” pois a sua estimação é feita de modo empírico.

Outra característica das médias móveis relaciona-se com o fato de todas as observações possuírem o mesmo peso. Mas, é razoável supor que as observações mais recentes devem possuir um peso maior. Assim, freqüentemente usa-se um cálculo ponderado para as médias móveis, com o objetivo de valorizar as últimas observações, em detrimento das mais antigas.

Mediana

A principal característica da mediana é dividir um conjunto ordenado em dois grupos iguais; metade terá valores inferiores e metade valores superiores à mediana. Em geral, a mediana ocupa a posição $(n + 1) / 2$ do conjunto ordenado, onde n corresponde à quantidade de elementos do grupo.

Para conjuntos com quantidade par de elementos a mediana é calculada através da média dos dois valores do meio, ou seja, a mediana fica entre as duas observações centrais da disposição ordenada. A mediana é a média aritmética dos valores numéricos correspondentes àquelas duas observações centrais.

Moda

Moda é o valor que ocorre com maior freqüência num conjunto, ou seja, é o valor que possui maior quantidade no conjunto. Pode ser obtida através de uma

disposição ordenada e, diferentemente da média aritmética, não é afetada pela ocorrência de quaisquer valores extremos (LEVINE et al, 2000, p. 123).

3.2.2. Medidas de Dispersão

Variância

A variância é uma medida de dispersão que serve para representar, com um único número, o quão próximos os valores observados em um conjunto de dados estão uns dos outros. Esta medida tem na média seu ponto de referência e o valor zero informa que não há dispersão no conjunto de observação.

A variância s^2 é definida como:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{Eq. (4)}$$

onde n corresponde ao número de dados do conjunto, \bar{x} = média aritmética e x_i = valor da observação.

Desvio padrão

Considera-se o desvio padrão como a mais importante medida de dispersão, tanto pelo seu extenso uso quanto por algumas de suas propriedades que estão descritas a seguir.

O desvio padrão é definido como a raiz quadrada da variância e é estimado pela equação:

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{Eq. (5)}$$

onde n corresponde ao número de dados do conjunto, \bar{x} = média aritmética e x_i = valor da observação.

A informação que o desvio padrão dá sobre a variabilidade deve ser entendida como a variabilidade que é apresentada relativamente a um ponto de referência - a média, e não propriamente a variabilidade dos dados, uns relativamente aos outros.

Se os dados se distribuem de forma normal, verifica-se que basicamente 68% dos dados estão no intervalo $[\bar{x} - s, \bar{x} + s]$, 95% no intervalo $[\bar{x} - 2s, \bar{x} + 2s]$ e 99% no intervalo $[\bar{x} - 3s, \bar{x} + 3s]$. Isto permite que se façam análises sobre o número de observações que caem longe da média aritmética, em termos de desvio padrão.

3.2.3. Coeficiente de Correlação

Um problema com o qual nos deparamos freqüentemente é o de estudar a correspondência existente entre duas variáveis, verificando se existe alguma relação entre elas e em que grau está presente. Um método simples e objetivo para tratar este problema é através do uso do coeficiente de correlação. Esta técnica aplica-se a dados bivariados (x_i, y_i) e verifica a existência de associações lineares entre as amostras (BUSSAB; MORETTIN, 2001, p. 63).

Uma característica importante deste método recai no fato de que o cálculo do coeficiente de correlação independe da unidade das medidas, ou seja, pode-se analisar a existência de relação entre medidas distintas, tais como peso (kg) e altura (m) de uma amostra populacional.

O coeficiente de correlação (linear) é representado por r e é calculado a partir da expressão:

$$r = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_x} \right) \left(\frac{y_i - \bar{y}}{\sigma_y} \right) \quad \text{Eq. (6)}$$

onde n = número de elementos da amostra;

x_i, y_i = elementos no ponto i ;

\bar{x}, \bar{y} = média aritmética dos elementos x e y , respectivamente; e

σ_x, σ_y = desvio padrão dos elementos x e y , respectivamente.

O coeficiente de correlação varia entre -1 e $+1$. O valor zero indica ausência de correlação entre as variáveis. Quanto maior for o módulo de r , maior será, o grau de associação linear existente entre as duas variáveis.

Quando $r > 0$, significa que a relação entre os elementos x e y é do mesmo sentido, isto é, valores grandes de x correspondem, de um modo geral, a valores grandes de y e vice-versa. Quando $r < 0$, a relação entre as variáveis x e y é inversa, significando que valores grandes de x , correspondem a valores pequenos de y e vice-versa.

Alta correlação, não implica, necessariamente, em causalidade (SINCICH, 1996). Se o coeficiente r for próximo de 1 ou de -1 é incorreto concluir que a troca de um elemento x corresponderá à troca em y . Pode-se, neste caso, concluir tão-somente a existência de tendência entre os elementos x e y .

3.3. TRANSFORMAÇÃO DE DADOS

Transformação de dados é um conjunto de técnicas para alterar ou consolidar os dados visando facilitar sua manipulação ou, até mesmo, reduzir o esforço computacional inerente ao processo.

Duas são as razões responsáveis pela transformação dos dados: a primeira quando é verificada a necessidade de se corrigir a existência de exceções ocorridas durante as suposições estatísticas e a segunda para melhorar o relacionamento entre as variáveis (HAIR; ANDERSON et al, 1998, p. 76-78).

As transformações podem ser realizadas em virtude de exigências teóricas, baseadas na natureza dos dados, ou em função das suas características. Na primeira têm-se as redes neurais como exemplo típico, aonde por questões de implementação os dados, geralmente, são normalizados durante a fase de pré-processamento. Já a segunda ocorre em decorrência da análise/exame do conjunto. Esta análise pode ser visual ou através do uso de técnicas que demonstrem a necessidade de se mudar alguns atributos dos dados, como, por exemplo, corrigir a sua distribuição normal.

Assim, segundo Han (HAN; KAMBER, 2001, p. 114-116), têm-se 5 formas de transformação:

- *Smoothing* (amortecimento): aonde se remove o ruído dos dados (variações de mais alta frequência), através do uso de regressão e média móvel;
- Agregação: Através da agregação pode-se reduzir a base agrupando-se os dados. Por exemplo: ao invés de se trabalhar com valores de vendas diárias pode-se usar os valores mensais, reduzindo a quantidade de informações a serem processadas;
- Generalização: a generalização compreende o aumento do nível hierárquico dos dados, ou seja, transformam-se informações primitivas

em conceito superior. Por exemplo, ao invés de se tratar o atributo idade, generaliza-se para o conceito de jovem, senhor e idoso.

- Construção de atributos ou características: aonde novos atributos são adicionados à base de dados para auxiliar o entendimento da estrutura multidimensional dos dados. Por exemplo: em uma base de dados existem informações sobre a largura, o comprimento e a altura de peças fabricadas. Pode tornar-se mais simples e conveniente usar o volume do sólido, ao invés de se trabalhar com as três medidas iniciais;
- Normalização: é uma das técnicas mais aplicadas em redes neurais e consiste na mudança de escala dos números para intervalos menores.

A transformação mais indicada para cada problema é aquela que apresenta a maior probabilidade de tornar a distribuição dos dados normal. Se a transformação utilizada não for a adequada, a normalidade pode vir a piorar, em vez de melhorar (CAMPOS, 2000/2001, cap. 13).

De todos os tipos de transformações apresentadas, a normalização dos dados é a técnica mais utilizada pois, em geral, os dados são muito sensíveis à diferença de escala ou de magnitude (HAIR; ANDERSON et al, 1998, p. 77). As normalizações geralmente são realizadas de tal forma que os valores fiquem situados na faixa entre 0 e 1 ou -1 e 1 . Entretanto, pode-se normalizar os dados para qualquer intervalo definido. Há normalizações lineares e não lineares.

Na categoria das normalizações lineares, encontramos a normalização min-max, aonde é aplicado sobre os dados a seguinte fórmula:

$$v' = \frac{v - \min}{\max - \min} (\max' - \min') + \min' \quad \text{Eq. (7)}$$

onde: v' = novo valor da variável

\min = menor valor existente na base;

\max = maior valor existente na base;

\min' = menor valor do novo intervalo de transformação;

\max' = maior valor do novo intervalo de transformação.

Entre os métodos de normalização não-linear, o mais usado é o Z-score que se baseia na média e desvio-padrão dos dados e é calculado através da fórmula:

$$v' = \frac{v - \bar{v}}{\sigma_v} \quad \text{Eq. (8)}$$

onde \bar{v} é a média do conjunto e σ_v corresponde ao desvio-padrão.

Um valor positivo para o Z-score indica que o resultado correspondente está acima da média. Já o negativo representa um resultado abaixo da média e o zero igual à média.

O uso de normalização dos dados facilita a comparação entre as variáveis, pois todas estarão na mesma escala. Dessa forma, os efeitos da escala são eliminados da base de dados.

3.4. VALORES INEXISTENTES (“MISSING VALUES”)

A inexistência ou falta de valores na base de dados está presente em grande parte das pesquisas e, principalmente no estudo de séries temporais, pode ser um problema crítico. Este fenômeno pode induzir a erros, comprometendo todo o esforço gasto, e muitas são as razões para a sua ocorrência, tais como: recusa em se prestar a informação, incapacidade do entrevistado, indisponibilidade do meio pesquisado, não retorno de questionários, impossibilidade de contato com a fonte, perda de registros, dentre outros. A ausência de informações é tratada, na literatura especializada, como erro e encontra-se no mesmo grupo dos erros de medidas e de processamento (SÄRNDAL; SWENSSON; WRETMAN, 1992. p. 551-556).

A inexistência de dados pode ocorrer por muitas razões e em várias situações sendo que os procedimentos a serem adotados podem ser classificados em uma de quatro categorias, baseados na aleatoriedade dos valores ausentes e no método usado para estimar os dados inexistentes (HAIR; ANDERSON et al, 1998, p. 51-53).

- O primeiro e mais simples procedimento a ser adotado nos casos de ausência de informação e também o mais utilizado é aquele que elimina os registros em que algum campo esteja ausente, ou seja, só se trabalha com registros completos. Deve-se observar se o conjunto final não terá o seu tamanho tão reduzido que prejudique as análises futuras;
- O segundo método consiste em se eliminar de todos os registros a variável que possua alto índice de inexistência na base de dados. Este

método, apesar de simples, deve ser adotado com cautela, cuidando-se para que a base não fique comprometida em função da inexistência de algumas variáveis;

- A terceira categoria abrange os métodos de estimação dos dados inexistentes baseado em valores válidos de outras variáveis.

Estimação dos dados é o processo de empregar conhecidas relações entre variáveis para identificar quantidades válidas a fim de se inferir os valores a serem inseridos na base de dados. Uma abordagem completa desses métodos, tais como: regressão, estimação múltipla, substituição etc, pode ser encontrada em (HAIR; ANDERSON et al, 1998);

- Por fim, a quarta e última categoria trata dos métodos utilizados para se incorporar dados no conjunto de registros. Esta técnica difere da anterior basicamente pelo fato de que a primeira trabalha com estimativas de valores e esta é basicamente estatística, ou seja, utiliza-se probabilidades, cálculos de correlações, médias e desvios padrões para gerar o valor faltante.

3.5. VALORES ESPÚRIOS (“*OUTLIERS*”)

Em um conjunto de dados pode-se encontrar determinadas observações que fogem ao comportamento geral da base como um todo. Em alguns casos apresentam valores grosseiramente diferentes ou inconsistentes com o restante do conjunto. Este grupo de dados é definido como *outliers* e a sua existência pode acarretar análises divergentes bem como inviabilizar todo o esforço de pesquisa.

Outliers podem ser considerados como ruídos ou exceções. Entretanto, em algumas aplicações o *outlier* é justamente o que se busca no estudo (HAN; KAMBER, 2001, p. 25). Como exemplo citamos a detecção de fraudes em cartão de crédito, onde a busca é justamente por dados que fogem ao padrão geral de uso do cartão. O que em outros casos é considerado ruído passível de ser descartado, nesta situação é o foco do estudo.

Segundo Hair (HAIR; ANDERSON et al, 1998, p. 64-67) existem 4 classes em que se podem classificar os *outliers*. Na primeira encontram-se os dados que decorrem de erros referentes aos procedimentos, ou seja, dados impostados por engano

ou mesmo regras não definidas corretamente nos algoritmos. Para este caso, os autores sugerem a eliminação dos registros.

Na segunda classe tem-se o resultado decorrente de um evento extraordinário, uma exceção que não necessariamente representa um erro. Nesta hipótese deve-se avaliar a conveniência de se permanecer ou não com os dados na base.

Na terceira classe inclui-se as observações para as quais não se tem uma explicação plausível. Deve-se separar estes dados a fim de se promover uma análise distinta desta informação, pois pode representar um segmento válido.

Por fim, na quarta e última classe encontram-se os valores que são combinações de outras variáveis. Nestas situações deve-se reter os dados que são combinações junto aos demais, a menos que evidências apontem para sua exclusão.

Os problemas comentados anteriormente não são triviais e, para isto, diversas abordagens do assunto já foram realizadas, tentando-se encontrar um método capaz de responder a questão.

Em uma primeira análise, a melhor forma de detecção ainda é a visual. Busca-se, através de gráficos, verificar quais pontos estão divergindo dos demais. Passada esta etapa, analisa-se, caso a caso, os pontos escolhidos como sendo candidatos a exclusão.

Além da análise gráfica, diversos outros métodos são utilizados para a detecção de *outliers*. Dentre estes os mais utilizados são a análise estatística dos dados e a categorização. No primeiro busca-se encontrar os pontos cujos valores estejam fora de um intervalo previamente determinado; já a segunda técnica procura separar em grupos os valores com características semelhantes.

3.6. ANÁLISE DO CAPÍTULO

As técnicas abordadas neste capítulo foram responsáveis por grande parte do esforço dedicado a este trabalho de pesquisa. O tratamento dos dados, aí compreendendo as normalizações, tratamento dos valores inexistentes e verificação dos *outliers* consumiram cerca de 20% do tempo. O restante foi distribuído entre a construção dos modelos (65%), avaliação dos resultados (10%) e comparações entre os métodos clássicos e não clássicos (5%). A seguir tem-se gráfico ilustrativo.

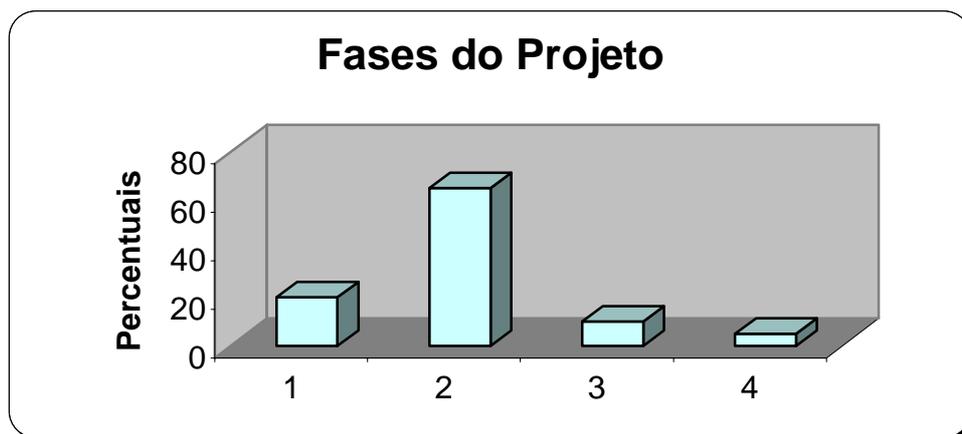


Figura 1: Figura representativa das fases do projeto: fase 1 => tratamento dos dados; fase 2 => Modelagem e treinamento da rede; fase3 => Avaliação do resultado e fase 4 => Comparação com modelos estatísticos.

A complexidade da série sob estudo tornou não trivial a construção e ajustes dos parâmetros do modelo neural. Ao contrário, após várias tentativas em se obter um bom desempenho nos treinamentos voltava-se para a fase de pré-processamento com o intuito de avaliar e utilizar novos métodos para tratamento dos dados visando facilitar o treinamento das redes.

Apesar da importância do tema, poucos autores ligados à construção de redes neurais artificiais dedicam espaço em seus livros para discorrer sobre técnicas de pré e pós-processamento, tendo que se recorrer a literatura de outras áreas, notadamente em estatística (LITTLE; RUBIN, 1987) (HAIR; ANDERSON et al, 1998) e *data mining* (HAN; KAMBER, 2001).

Segundo Bishop (BISHOP, 1996, p. 295): *“Outro caminho através do qual a performance da rede pode ser melhorada, algumas vezes dramaticamente, é incorporando conhecimento prévio (prior knowledge) aos dados. Tal conhecimento se refere a relevantes informações que podem ser utilizadas no desenvolvimento da solução, melhorando o treinamento dos dados.”*

Vê-se com isto que apesar de trabalhosa a modelagem dos dados pode ser responsável por um melhor ou pior desempenho da rede. No capítulo seguinte apresenta-se a análise das séries em estudo através das técnicas de pré-processamento nos dados.

4. ANÁLISE DAS SÉRIES EM ESTUDO

Neste capítulo apresenta-se, no primeiro item, o tratamento que foi adotado para os registros inexistentes na base de dados, em seguida discorre-se sobre o trabalho realizado nos registros espúrios para, na última seção, abordar as técnicas de normalização de dados que foram empregadas.

4.1. TRATAMENTO DOS DADOS

Na análise do problema, descrita no capítulo 2, a hipótese onde todas as agências são receptoras nas segundas e terças-feiras; equilibradas nas quartas-feiras e pagadoras nas quintas e sextas-feiras não se sustenta com a realidade. Mais ainda: este modelo não é válido nem mesmo para uma única agência.

Em razão do exposto, procurou-se detectar a existência de algum padrão no movimento de saques e depósitos nas agências, ou seja, se os dias em que o movimento da agência é positivo ou negativo têm relação com algum outro intervalo em período posterior. Objetiva-se com isto descobrir a existência de ciclos bem definidos de movimentos a fim de se adotar este parâmetro como a janela ideal para entrada de dados na rede.

Para realizar este estudo substituíram-se os vetores originais por outros onde, por convenção, os dias em que a variação do movimento foi igual ou maior que zero (depósitos) foram considerados como 1 e os dias em que a variação do movimento foi negativa (saques) foram considerados como 0. Posteriormente, se dividiu o vetor em períodos de tempo anual para cada agência. Assim, cada agência – 02, 06 e 13 – ficou com 4 vetores correspondentes aos anos de 1998, 1999, 2000 e 2001. Cabe ressaltar que o vetor de 2001 possui apenas 3 meses de registros.

Em seguida duplicou-se cada um dos vetores, fixando o primeiro e deslocando o segundo de 1 a 200 posições (no caso do ano de 2001 o deslocamento foi de 1 a 60 posições). A cada deslocamento promovido, calculou-se o percentual de valores coincidentes entre o primeiro e o segundo vetor, ou seja, a quantidade de 0's e 1's que coincidiriam. Abaixo tem-se um exemplo para ilustrar a análise realizada.

Quadro 14: Análise de ciclicidade da série. Vê-se que no deslocamento de 1 posição há coincidência em 5 registros que estão sombreados.

1	1	1	0	0	0	1	0	1	0	1	1	0	0	1	1	1	0	1	0	
	1	1	1	0	0	0	1	0	1	0	1	1	0	0	1	1	1	0	1	0

Verificou-se, durante o trabalho, que o deslocamento de 5 posições foi o que mais se destacou globalmente, coincidindo, os valores positivos e negativos, em aproximadamente 70% das ocorrências, apesar de haver tido anos em que deslocamentos de 20, 23 ou mesmo 45 posições tenham alcançado percentuais maiores de coincidência entre os vetores.

Estes resultados (70% de coincidência para deslocamento de 5 posições) situaram-se em patamares relativamente baixos para a expectativa e atribuiu-se tal performance ao fato da ausência de valores na base de dados em razão de feriados, pois esta situação pode acarretar distorções na análise assim como prejudicar os resultados em função do potencial escondido nas informações inexistentes (HAIR; ANDERSON et al, 1998, p. 47).

Avaliando os resultados verificou-se que a comparação entre semanas onde exista um feriado no intervalo sofrerá um deslocamento do vetor de dados fazendo com a comparação se dê entre dias diferentes de semana, ou seja, se o deslocamento for de cinco posições, correspondentes a uma semana, a partir do momento da existência de um feriado os dias da semana não serão mais coincidentes, pelo menos até se alcançar algum múltiplo de 5 feriados. Assim, optou-se por incluir registros nos dias faltantes de tal forma que os dias da semana ficassem sempre completos. Acreditava-se que com esta atitude os índices de acerto sofreriam significativa melhora.

No caso específico desta dissertação em que apenas uma única variável é o objeto de estudo as técnicas de relacionamento com outras componentes (correlação, regressão etc) não podem ser aplicadas para o cálculo da variável faltante. A interpolação de dados também não é ideal para este caso, pois o dado interpolado se situará próximo dos antecessores e sucessores. Assim, feriados na segunda-feira onde na maioria das vezes é de recebimento sofreria a influência da sexta-feira e poderia ser caracterizado como dia de pagamento.

Optou-se, então, por realizar 3 análises distintas, onde a primeira seria com a inserção de registros 0 para os dias faltantes, caracterizando todas as ausências como sendo de movimento pagador. Na segunda análise inseriu-se registros 1 para os dias

faltantes caracterizando-os como de movimento receptor e, por último, incluiu-se o valor 99 nos dias inexistentes e quando da comparação dos vetores não se computou estes dias para efeito de cálculos.

A seguir têm-se os resultados referentes ao caso 1, onde se completou com 0 os registros faltantes, para as três agências.

Os quadros abaixo apresentam na linha “Mínimo” os menores percentuais de coincidência de registros e na linha “Janela mínima” o tamanho do deslocamento que gerou este percentual. Nas linhas subsequentes (“Máximo”, “Janela máxima” e “Reg. Certos”) são apresentados os valores correspondentes aos acertos, tamanho da janela e quantidade de registros que coincidiram. Cabe destacar que na coluna da direita destas mesmas linhas foi colocado, para comparação, os resultados utilizando-se a janela de 5 posições pois, como foi comentado, foi este tamanho de janela que, no geral, apresentou os melhores resultados.

Quadro 15: Agência 02 – Comparação dos dias com base completada com valor positivo.

	Agência 02 (série completada com valor positivo nos registros faltantes)							
	Ano							
	1998 (259 reg)		1999 (261 reg)		2000 (260 reg)		2001 (65 reg)	
Mínimo	39,77%		40,23%		41,54%		60,61%	
Janela mínima	32		97		54		19	
Máximo	69,88%	63,71%	72,41%	64,75%	70,77%	64,62%	78,79%	72,72
Janela máxima	20	5	20	5	65	5	1	5
Reg. certos	181	165	189	169	184	168	52	48

Quadro 16: Agência 06 – Comparação dos dias com base completada com valor positivo. Não há informação nas áreas sombreadas em função de a janela de 5 posições haver tido o melhor desempenho.

	Agência 06 (série completada com valor positivo nos registros faltantes)							
	Ano							
	1998 (259 reg)		1999 (261 reg)		2000 (260 reg)		2001 (65 reg)	
Mínimo	45,95%		44,83%		47,69%		56,92%	
Janela mínima	122		72		8		2	
Máximo	72,29%		77,01%		73,08%		96,92%	
Janela máxima	5		5		5		5	
Reg. Certos	195		201		190		63	

Quadro 17: Agência 13 – Comparação dos dias com base completada com valor positivo.

	Agência 13 (série completada com valor positivo nos registros faltantes)							
	Ano							
	1998 (259 reg)		1999 (261 reg)		2000 (260 reg)		2001 (65 reg)	
Mínimo	59,85%		54,02%		46,15%		47,69%	
Janela mínima	36		76		33		14	
Máximo	76,06%	66,80%	74,71%	65,52%	64,62%	61,54%	69,23%	56,93
Janela máxima	22	5	1	5	25	5	20	5
Reg. certos	197	2	195	171	168	2	45	37

A seguir tem-se o resultado referente à 2ª. análise, onde os registros faltantes foram completados com 1.

Quadro 18: Agência 02 – Comparação dos dias com base completada com valor negativo.

	Agência 02 (série completada com valor negativo nos registros faltantes)							
	Ano							
	1998 (259 reg)		1999 (261 reg)		2000 (260 reg)		2001 (65 reg)	
Mínimo	42,08%		41,00%		43,08%		53,85%	
Janela mínima	32		79		77		13	
Máximo	68,34%	62,16%	73,95%	66,28%	71,54%	66,15%	72,31%	66,15
Janela máxima	20	5	20	5	65	5	1	5
Reg. certos	177	161	193	173	186	172	47	43

Quadro 19: Agência 06 – Comparação dos dias com base completada com valor negativo.

	Agência 06 (série completada com valor negativo nos registros faltantes)							
	Ano							
	1998 (259 reg)		1999 (261 reg)		2000 (260 reg)		2001 (65 reg)	
Mínimo	41,31%		44,06%		41,54%		50,77%	
Janela mínima	122		17		8		7	
Máximo	71,43%		72,41%		70,77%		90,77%	87,69
Janela máxima	5		5		5		25	5
Reg. certos	185		189		184		59	57

Quadro 20: Agência 13 – Comparação dos dias com base completada com valor negativo.

	Agência 13 (série completada com valor negativo nos registros faltantes)							
	Ano							
	1998 (259 reg)		1999 (261 reg)		2000 (260 reg)		2001 (65 reg)	
Mínimo	52,90%		49,43%		43,08%		44,62%	
Janela mínima	36		116		12		11	
Máximo	73,75%	63,71%	71,65%	58,62%	63,85%	58,46%	63,08%	56,92
Janela máxima	22	5	66	5	130	5	4	5
Reg. certos	191	165	187	153	166	152	41	37

Por fim, abaixo estão os quadros em que foi adotado o conceito de se completar a série com registros representativos dos valores faltantes mas, quando da comparação, não se computou os dias em que não havia movimento. A inserção dos valores serviu, tão-somente, para que todas as semanas ficassem com 5 dias úteis e assim facilitasse a comparação entre semanas.

Quadro 21: Agência 02 – Comparação dos dias com a base completa.

	Agência 02 (série completada mas os registros não foram computados)							
	Ano							
	1998 (259 reg)		1999 (261 reg)		2000 (260 reg)		2001 (65 reg)	
Mínimo	36,68%		36,40%		39,62%		52,31%	
Janela mínima	32		97		99		19	
Máximo	64,86%	58,69%	68,97%	61,30%	67,31%	61,54%	70,77%	64,62
Janela máxima	20	5	20	5	65	5	1	5
Reg. certos	168	152	180	160	175	160	46	42

Quadro 22: Agência 06 – Comparação dos dias com a base completa.

	Agência 06 (série completada mas os registros não foram computados)							
	Ano							
	1998 (259 reg)		1999 (261 reg)		2000 (260 reg)		2001 (65 reg)	
Mínimo	39,38%		40,23%		40,77%		49,23%	
Janela mínima	122		72		8		7	
Máximo	69,11%		70,50%		68,08%		89,23%	87,69
Janela máxima	5		5		5		25	5
Reg. certos	179		184		177		58	57

Quadro 23: Agência 13 – Comparação dos dias com a base completa.

	Agência 13 (série completada mas os registros não foram computados)							
	Ano							
	1998 (259 reg)		1999 (261 reg)		2000 (260 reg)		2001 (65 reg)	
Mínimo	52,12%		47,89%		40,77%		41,54%	
Janela mínima	36		76		53		14	
Máximo	70,66%	61,00%	68,97%	57,85%	60,00%	56,15%	61,54%	52,31
Janela máxima	22	5	108	5	25	5	20	5
Reg. certos	183	158	180	151	156	146	40	34

Apesar da janela de 5 dias não ter sido a melhor em todos os experimentos, os valores obtidos nesta situação foram, no geral, sempre próximos do melhor resultado. Além disso, não se descobriu nenhum padrão para as janelas diferentes de 5. Assim, optou-se por realizar a análise dos resultados comparando-se sempre com a janela de 5 dias úteis.

Outro fato que chamou a atenção foi a agência 06 ter sido a mais homogênea de todas, tendo os seus melhores resultados sido obtidos com uma janela de 5 dias úteis. Esta agência é a que possui o maior desvio padrão entre todas elas e foi caracterizada como tendo frequência de oscilação alta (vide análise preliminar), entretanto o comportamento dela é bem definido e, por isto, além de ter tido os melhores resultados com uma janela de 5 dias úteis teve também, na média, os melhores percentuais de acertos.

A agência 13 que era considerada como a de melhor comportamento apresentou, na média, os piores resultados. Ou seja, as previsões baseadas somente na verificação de o movimento ser positivo ou negativo desta agência foram piores que as demais agências.

Algumas conclusões podem ser tiradas deste experimento: a melhor janela para se trabalhar seria aquela cujos valores correspondem a 5 dias úteis. Isto demonstra uma certa lógica pois esta janela corresponde a uma semana e é razoável supor que haja um certo comportamento cíclico neste intervalo. Entretanto, também seria razoável

imaginar que períodos mensais, semestrais etc, também apresentassem o mesmo tipo de comportamento. Isto não foi verificado. Teve-se bons resultados para períodos de 20, 22 e até mesmo 25 dias e isto corresponde a aproximadamente 1 mês (dias úteis). Mas, no geral, poucos foram os anos e agências que tiveram estes últimos períodos como os melhores. Isto, talvez, seja decorrência do fato de cada mês possuir uma quantidade de dias úteis distinta e é extremamente trabalhoso e complexo preparar o complemento desses meses de tal forma que todos eles tenham a mesma quantidade de dias.

O fato de se completar as semanas melhorou os resultados obtidos mas de um percentual muito pequeno. Não foi observado nenhum grande avanço usando-se desta técnica e, por este motivo, resolveu-se adotar, sem completar os registros faltantes, a janela de 5 posições como a ideal.

4.2. TRATAMENTO DOS OUTLIERS

Verificou-se, nas análises realizadas, a presença de valores extremamente altos/baixos que poderia vir a dificultar o aprendizado da rede. Dessa forma, estudos foram feitos visando minimizar ou, até mesmo, eliminar os efeitos oriundos destes dados. É importante lembrar que estes registros não são ruídos existentes na série e sim uma característica da base de dados. Todos estes pontos foram analisados individualmente.

Primeiramente pesquisou-se a quantidade de pontos existentes que extrapolavam parâmetros pré-definidos de 1 e 2 desvios padrões. Com isto objetivou-se encontrar o ponto de corte ideal para considerar registros como sendo *outliers*. O resultado encontra-se no quadro abaixo, onde se tem a quantidade de registros que ficaram abaixo de 1 e 2 desvios padrões (colunas DP e DP*2, respectivamente), bem como a quantidade de dados acima destes limites, representados nas colunas “*outliers*”.

Quadro 24: *Outliers encontrados nas séries de dados das agências 02, 06 e 13.*

	DP	Outliers	%	DP * 2	Outliers	%
Ag. 02	612	199	75,46	769	42	94,82
Ag. 06	647	164	79,78	769	42	94,82
Ag. 13	662	149	81,63	771	40	95,07

Vê-se que o parâmetro estipulado de um desvio padrão e dois desvios padrões mantiveram uniformidade de valores, ou seja, as quantidades de pontos nesta situação são praticamente iguais para as três agências. Além disto, o percentual de

valores válidos, ou seja, aqueles dentro do limite estipulado de 2 desvios padrões é bem razoável, com 94,82% para as agências 02 e 06 e 95,07% para a agência 13.

Esta dissertação trabalhará, em função da análise realizada, com o parâmetro de dois desvios padrões. Em alguns experimentos os valores fora do intervalo de 2DP serão trazidos para este patamar e, em outros casos, o experimento será realizado com a série completa, sem nenhuma modificação. Em todos os casos será informado anteriormente qual o critério adotado.

O estudo dos *outliers* é uma importante etapa na análise dos dados, pois a existência de valores divergentes do conjunto pode comprometer ou desvirtuar a análise que está sendo realizada. Assim, após a definição dos valores considerados como fora do padrão da série (superiores a dois desvios padrões), efetuou-se um levantamento destes dados, ou seja, tentou-se classificar estes registros em função das suas características.

Após a checagem dos dados e verificado que não houve erro de transcrição ou de captura dos registros, optou-se por analisar o contexto em que estes valores estão inseridos, ou seja, baseado na data da ocorrência procurou-se descobrir algum acontecimento que pudesse interferir de forma tão significativa na normalidade da série. Como os eventos mais evidentes onde há ocorrência de forte movimento bancário são os feriados, estudou-se a relação entre os feriados nacionais e as ocorrências verificadas.

Quadro 25: Quadro resumo dos feriados nacionais no período sob estudo.

	Data	1998	1999	2000	2001
Confraternização Universal	01.01	quinta-feira	sexta-feira	sábado	segunda-feira
Carnaval		segunda-feira (23/02) terça-feira (24/02)	segunda-feira (15/02) terça-feira (16/02)	segunda-feira (06/03) terça-feira (07/03)	segunda-feira (26/02) terça-feira (27/02)
Semana Santa		quinta-feira (09/04) sexta-feira (10/04)	quinta-feira (01/04) sexta-feira (02/04)	sexta-feira (21/04) junto com Tiradentes	sexta-feira (13/04)
Tiradentes	21.04	terça-feira	quarta-feira	sexta-feira	sábado
Trabalho	01.05	sexta-feira	sábado	segunda-feira	terça-feira
Corpus Christi		quinta-feira (11/06)	quinta-feira (03/06)	quinta-feira (22/06)	quinta-feira (14/06)
Independência	07.09	segunda-feira	terça-feira	quinta-feira	sexta-feira
N.S.Aparecida	12.10	segunda-feira	terça-feira	quinta-feira	sexta-feira
Finados	02.11	segunda-feira	terça-feira	quinta-feira	sexta-feira
Proclamação	15.11	domingo	segunda-feira	quarta-feira	quinta-feira
Natal	25.12	sexta-feira	sábado	segunda-feira	terça-feira

No item 1 do Anexo têm-se, em ordem cronológica, os registros espúrios das três dependências, com a última coluna descrevendo o fato que pode ter sido responsável pela inconsistência. Nos casos em que não se detectou nenhuma ocorrência passível de ser responsável pela origem do dado espúrio foi posta a expressão “Sem justificativa”.

Os estudos visaram descobrir um padrão de comportamento para os *outliers* (valores superiores a dois desvios padrões). As três agências juntas somam 124 registros considerados *outliers* sendo que 74 (59,68%) desses registros não tiveram justificativa aparente, tais como data próxima a feriado, boatos econômicos etc. Os demais são referentes ou à véspera de feriados ou início da semana mas, mesmo assim, o que vale para uma agência não necessariamente vale para as outras. Verificou-se, também, que 46 registros (37%) estão nos meses de janeiro e fevereiro.

Nenhuma data cujo valor tenha sido considerado espúrio apresentou o mesmo comportamento nas 3 agências simultaneamente, ou seja, o máximo encontrado foram duas agências com *outliers* no mesmo dia. Além disso, nos 3 anos e 3 meses de registros não se verificou nenhum feriado que tivesse gerado *outlier* para duas agências em toda a série, ou seja, que o mesmo feriado e as mesmas agências repetissem essa característica em todos os anos sob estudo. Abaixo apresenta-se os grupos de agências cujas datas dos *outliers* são coincidentes.

Quadro 26: Relação de feriados geradores de outliers para duas agências simultaneamente.

Agências	Data	Feriado
Agências 02 e 06	25.02.1998	quarta-feira de cinzas
	17.02.1999	quarta-feira de cinzas
	02.03.2000	antevéspera de carnaval
	03.03.2000	véspera de carnaval
Agências 02 e 13	14.04.1998	após semana Santa
	03.01.2000	primeiro dia útil do ano
	02.01.2001	primeiro dia útil do ano
Agência 06 e 13	20.04.2000	véspera de semana Santa

A distribuição dos *outliers* da agência 02 está equilibrada, com uma leve tendência para os meses de janeiro e fevereiro. Março também teve um dos maiores números de *outliers* mas estes pontos decorrem do fato de ter tido carnaval no período, época de maior movimento bancário. Dos 42 pontos analisados, 11 deles (26,19%) estão no período de janeiro e fevereiro, conforme pode ser visto no item 2 do Anexo.

Dos 42 pontos considerados *outliers* para a agência 06 um total de 26 desses pontos estão no período de janeiro e fevereiro, perfazendo 61,90%. Os outros 16 pontos estão espalhados pelos demais meses. A distribuição dos pontos está mostrada no item 2 do Anexo.

Por último, dos 40 pontos de *outliers* da agência 13, 9 deles estão no período de janeiro e fevereiro (22,5%). A agência 13 caracteriza-se pelo equilíbrio com maior concentração de *outliers* no mês de dezembro. Este é um mês realmente atípico em função do aquecimento da economia que ocorre no período – pagamento de 13º. salário e compras de natal. Os pontos estão apresentados no item 2 do Anexo.

A fim de verificar se existe algum padrão de comportamento dos *outliers* em relação ao dia útil do mês, assim como relativamente ao dia da semana (segunda-feira, terça-feira etc.) foi realizado um levantamento dos *outliers* com o foco nestes parâmetros. Primeiramente, foi analisado o relacionamento entre o dia útil do mês e a ocorrência de *outliers*. O resultado, por agência, encontram-se nos quadros existentes no item 2, do Anexo.

Os cinco primeiros dias úteis do mês são considerados como período de fortes pagamentos, haja vista ser a época em que se movimenta muito dinheiro referente a salários. Entretanto, pode-se deduzir dos quadros acima que a ocorrência de *outliers* neste intervalo não chega a definir um padrão único para todas as agências. Avaliando-se as ocorrências por agência vê-se que outros períodos possuem quantidades de *outliers* iguais ou até mesmo maiores que os 5 primeiros dias úteis do mês.

A agência 02, por exemplo, possui 14 dias com *outliers* no período do 8º ao 12º dia útil. Isto representa 33,33% dos dados, pouca coisa menor que os 42,86% (18 dias) relativos aos 5 primeiros dias úteis. Já a agência 06 detem no intervalo do 10º ao 14º dia útil 14 registros espúrios, superior, portanto, aos 12 registros encontrados nos 5 primeiros dias úteis do mês. Por fim, temos a agência 13 com 18 pontos no intervalo do 17º ao 21º dias úteis, bem superior aos 13 verificados nos 5 primeiros dias do mês.

A seguir tem-se um quadro resumo das quantidades de *outliers* por dia útil, separado por agência e por quantidade de movimentos pagadores e recebedores.

Vê-se que, olhando individualmente as agências, não se pode afirmar que haja um padrão típico de oscilações do movimento no início do mês. Nem mesmo um

padrão de fortes pagamentos foi verificado nos 5 primeiros dias úteis do mês pois dos 18 pontos ocorridos neste período na agência 02 em 6 deles os *outliers* foram referentes a recebimentos. Na agência 06 verificou-se 2 *outliers* de pagamentos e 10 de recebimentos e a agência 13 apresentou 11 de pagamentos e 3 de recebimentos.

Quadro 27: Resumo da quantidade de outliers por agência. A coluna "P" indica a quantidade de outliers referentes a movimentos de pagamentos e a coluna "R" de recebimentos.

Dia útil	Quantidade de outliers								
	Ag. 02	P	R	Ag. 06	P	R	Ag. 13	P	R
1º.	4	0	4	1	0	1	3	0	3
2º.	8	8	0	5	1	4	5	5	0
3º.	3	3	0	3	1	2	2	2	0
4º.	2	1	1	2	0	2	1	1	0
5º.	1	0	1	1	0	1	2	2	0
6º.	0	0	0	0	0	0	1	1	0
7º.	1	0	1	3	0	3	1	0	1
8º.	3	0	3	2	0	2	0	0	0
9º.	1	1	0	0	0	0	0	0	0
10º.	4	1	3	2	2	0	0	0	0
11º.	4	0	4	2	0	2	0	0	0
12º.	2	0	2	5	0	5	0	0	0
	Quantidade de outliers								
Dia útil	Ag. 02	P	R	Ag. 06	P	R	Ag. 13	P	R
13º.	0	0	0	2	0	2	0	0	0
14º.	0	0	0	3	1	2	2	2	0
15º.	1	1	0	1	0	1	1	1	0
16º.	1	0	1	2	0	2	2	2	0
17º.	1	0	1	4	1	3	1	0	1
18º.	3	2	1	2	0	2	5	4	1
19º.	3	3	0	2	2	0	8	6	2
20º.	0	0	0	0	0	0	2	2	0
21º.	0	0	0	0	0	0	2	2	0
22º.	0	0	0	0	0	0	1	1	0
23º.	0	0	0	0	0	0	1	1	0
Totais	42	20	22	42	8	34	40	32	8

Dando prosseguimento à análise de *outliers* buscou-se verificar a existência de padrões entre os dias da semana. A relação entre estas ocorrências para a agência 02 pode ser verificada no item 2, do Anexo.

Verifica-se que todos os movimentos de *outliers* nas segundas-feiras da agência 02 são de recebimentos assim como todos os movimentos das sextas-feiras são de pagamento, corroborando com o conceito de que o início da semana é caracterizado por entrada de numerário e o fim de semana de saída de dinheiro. Vê-se, também, que as ocorrências durante o meio da semana estão bem equilibradas, com 10 eventos de pagamento e 9 de recebimentos. No item 2, do Anexo, têm-se o quadro do movimento da agência 06.

Os *outliers* da agência 06 estão caracterizados como fortemente recebedores haja vista que dos 42 pontos considerados, 34 deles (80,95%) são referentes a recebimentos. Dois fatos chamam a atenção: o primeiro é que nas terças-feiras e quartas-feiras não existe nenhum pagamento, sendo todos os registros referentes a recebimentos; o outro diz respeito a ocorrências de recebimentos nas sextas-feiras, dia notadamente de pagamentos. Por fim tem-se a análise do movimento da agência 13 e o respectivo quadro dos *outliers* encontra-se no item 2, do Anexo.

Pode-se verificar que o movimento da agência 13 também foge ao padrão esperado, pois se tem um equilíbrio entre pagamentos e recebimentos nas segundas-feiras (3 pontos de pagamentos e 4 de recebimentos) o que não condiz com a expectativa. Além disso, nas sextas-feiras encontramos um recebimento que apesar de ser insignificante no conjunto (um registro de recebimento para dez de pagamentos) não o é no valor. Finalmente, o meio da semana não está devidamente caracterizado como equilibrado, onde encontramos muito mais pagamentos que recebimentos (86,36% de pagamentos e 13,64% de recebimentos). Adiante, tem-se quadro resumo dos *outliers* segundo os dias da semana.

Quadro 28: Resumo da quantidade de outliers por agência e dia da semana. A coluna "P" indica a quantidade de outliers referentes a movimentos de pagamentos e a coluna "R" de recebimentos.

	Quantidade de outliers								
	Ag. 02	P	R	Ag. 06	P	R	Ag. 13	P	R
Segunda-feira	12	0	12	5	0	5	7	3	4
Terça-feira	7	2	5	18	0	18	6	4	2
Quarta-feira	5	1	4	6	0	6	9	8	1
Quinta-feira	7	7	0	6	3	3	7	7	0
Sexta-feira	11	11	0	7	5	2	11	10	1
Totais	42	21	21	42	8	34	40	32	8

No quadro acima, analisando a quantidade total de *outliers* a agência 02 mantém equilíbrio, com 21 *outliers* de pagamento e 21 de recebimentos. Já a agência 06 possui uma forte tendência para exceções de recebimento. Por outro lado, a agência 13 é fortemente voltada para *outliers* de pagamentos.

Aqui cabe ressaltar que a agência 13 foi caracterizada pelo seu movimento como sendo uma agência recebedora pois em 70,65% dos dias ela teve como resultante movimentos de depósitos. Entretanto, a grande maioria dos *outliers* desta agência são de pagamento, o que, a princípio parece ser uma contradição. Acredita-se que tal fato é decorrente da região em que se insere a agência, ou seja, por se tratar de área agropecuária há, em determinadas épocas do ano, a necessidade de se retirar quantias vultosas para pagamento de trabalhadores temporários durante a fase de colheita dos produtos.

Fechando o tópico relativo à análise dos *outliers* pesquisou-se o relacionamento entre esses pontos e os meses do ano, a fim de verificar se há algum padrão. Acredita-se que os meses de janeiro, fevereiro e dezembro são de movimento muito grande, notadamente em dezembro com o advento do 13º. salário.

Quadro 29: Resumo da quantidade de outliers por agência e mês do ano. A coluna "P" indica a quantidade de outliers referentes a movimentos de pagamentos e a coluna "R" de recebimentos.

	Quantidade de outliers								
	Ag. 02	P	R	Ag. 06	P	R	Ag. 13	P	R
Janeiro	6	3	3	12	0	12	5	3	2
Fevereiro	5	1	4	14	3	11	4	3	1
Março	5	3	2	6	2	4	2	2	0
Abril	3	2	1	3	2	1	4	3	1
Maio	2	0	2	0	0	0	2	2	0
Junho	4	3	1	0	0	0	3	3	0
Julho	2	1	1	0	0	0	2	2	0
Agosto	3	2	1	0	0	0	1	1	0
Setembro	5	4	1	2	0	2	2	2	0
Outubro	3	0	3	2	0	2	2	2	0
Novembro	3	2	1	2	0	2	3	3	0
Dezembro	1	0	1	1	1	0	10	6	4
Totais	42	21	21	42	8	34	40	32	8

Não se conseguiu detectar padrão único para todas as agências. Enquanto na agência 02 se verifica um certo equilíbrio de ocorrências de outliers durante o ano, com pequeno crescimento nos meses de janeiro, fevereiro e março, a agência 06 mostra-se bem menos previsível com quase todas as ocorrências verificadas nos 4 primeiros meses do ano (83,33%). A agência 13 mantém equilíbrio durante os 11 primeiros meses do ano, e forte movimento em dezembro, correspondendo, individualmente, a 25% das ocorrências.

Após este conjunto de análises, pode-se afirmar que não existem padrões típicos de comportamento para as agências pesquisadas, nem por dia da semana, por dia do mês ou por mês do ano.

O Banco objeto do estudo possui aproximadamente 8.000 pontos de atendimento no país e, numa visão macro, incluindo o conjunto de todas as dependências, as hipóteses abaixo são válidas, conforme explicado pelos analistas da Empresa. Os cálculos de previsão do Banco para planejamento das suas atividades no dia seguinte ($d + 1$) são realizados baseados nestas premissas;

- início da semana com fortes recebimentos, equilíbrio no meio da semana e fortes pagamentos no fim da semana;
- fortes pagamentos nos primeiros dias úteis do mês com tendência ao equilíbrio nos dois últimos decêndios;
- saques vultosos nos meses de férias (janeiro e fevereiro) e de pagamento de 13º. salário (dezembro), com equilíbrio nos demais meses do ano.

Entretanto, estas mesmas hipóteses aplicadas em um universo micro, ou seja, por agência individualmente, não se confirmaram, conforme pode ser verificado nas análises efetuadas. À vista do exposto, três alternativas foram consideradas para se realizar um tratamento dos *outliers* existentes na série: 1^a. retirar os *outliers*; 2^a. interpolar os dados e 3^a. nivelar pelo teto de 2 desvios padrões.

Conforme já comentado, os *outliers* presentes na série não são ruídos e sim característica da base de dados. Assim, a retirada destes pontos distorceria a série em função do vácuo que ficaria no dia relativo ao *outlier*. Além disso, há de se considerar que o valor referente à data em que ocorreu o *outlier* também deve ser previsto, ou seja, com a retirada deste ponto dificulta-se a previsão para aquele dia.

A 2^a. hipótese que se apresenta – interpolação – também se mostra inadequada para o problema, haja vista que os valores interpolados serão necessariamente inferiores ao teto, o que geraria distorção.

Por fim, o nivelamento dos *outliers* pelo teto de 2 desvios padrões é o que se mostra mais adequado pois ao mesmo tempo em que não se exclui estes dados, têm-se valores suficientemente grandes para se apresentar à rede.

Abaixo encontram-se gráficos para cada agência, onde os *outliers* foram trazidos para o teto de 2 desvios padrões. A fim de melhor identificá-los, marcou-se os pontos originais com o símbolo “x”.

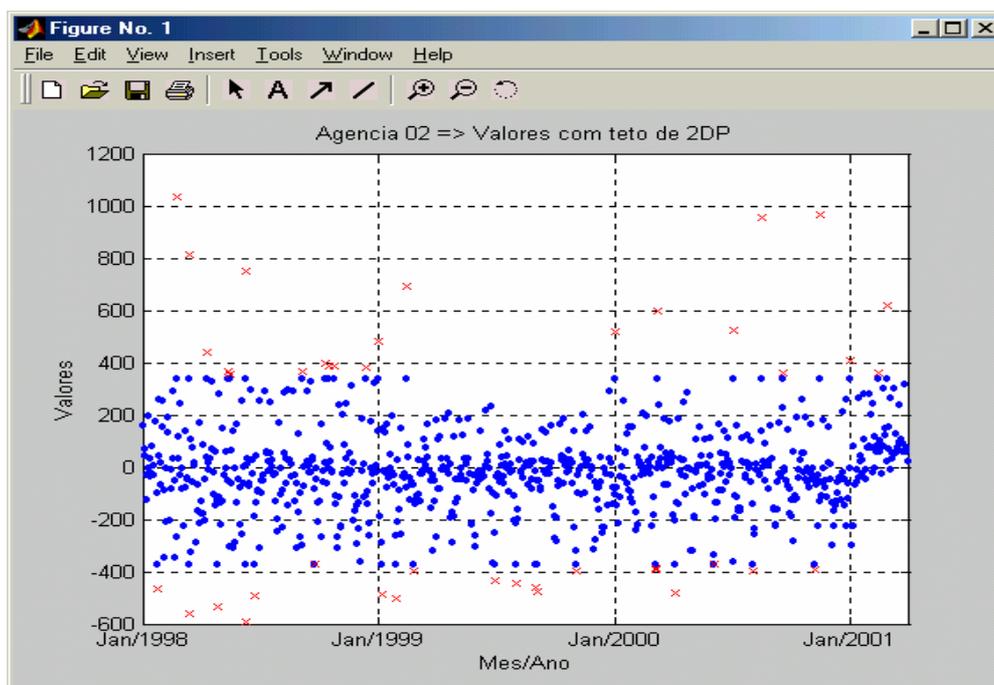


Gráfico 16: Ag. 02 – Gráfico com os pontos no teto de 2 desvios padrões e com marcação dos outliers.

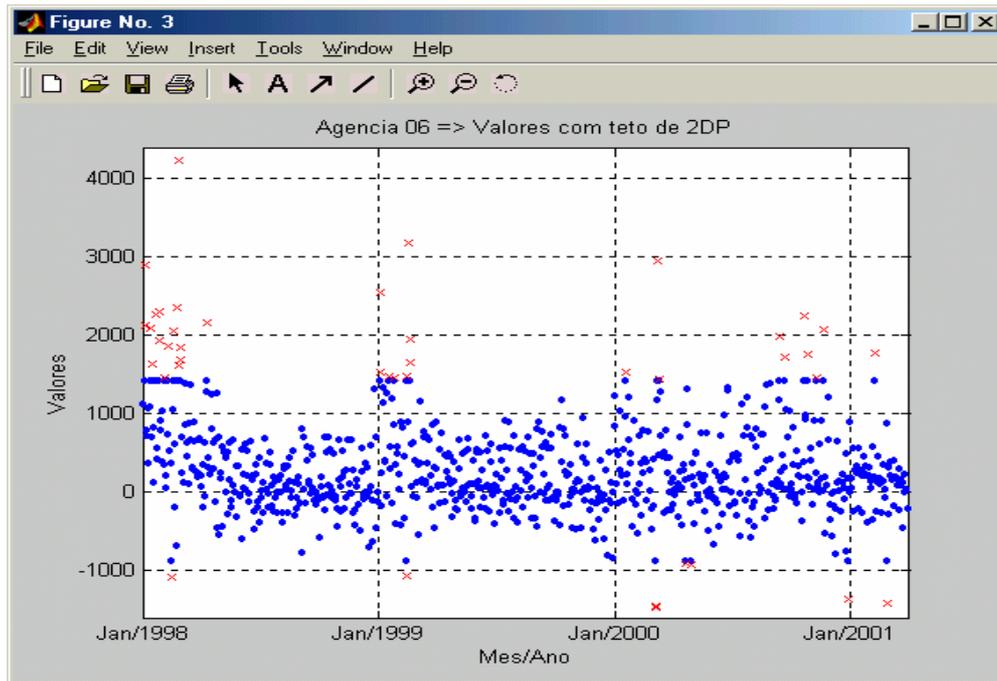


Gráfico 17: Ag. 06 – Gráfico com os pontos no teto de 2 desvios padrões e com marcação dos outliers.

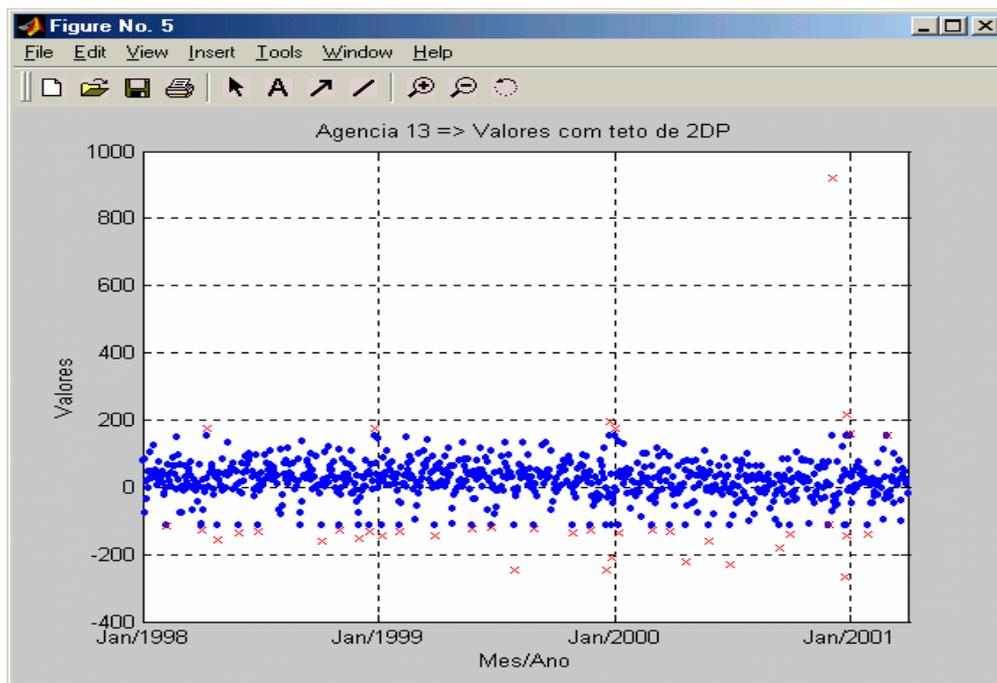


Gráfico 18: Ag. 13 – Gráfico com os pontos no teto de 2 desvios padrões e com marcação dos outliers.

4.3. NORMALIZAÇÃO DA BASE DE DADOS

Durante a pesquisa foram avaliados diversos modelos de normalização, a fim de se achar aquele que melhor se adequasse ao problema.

Inicialmente realizaram-se transformações lineares aplicando-se sobre os registros as seguintes fórmulas:

- Normalização 1 = $\frac{x}{\text{abs}(x_{\text{máx}})}$;
- Normalização 2 = $\frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$;
- Normalização 3 = $\left[\left(\frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \right) - 0.5 \right] * 2$;
- Normalização 4 = $\frac{x - \bar{x}}{\sigma}$;
- Normalização 5 = $\frac{x_{\text{atual}}}{x_{\text{anterior}}} - 1$.

Os dados após passarem por estes tipos de normalização foram aplicados à uma rede neural a fim de se testar o desempenho. Os resultados não foram satisfatórios, tendo ocorrido um acerto em aproximadamente 10% dos registros.

Avaliando os motivos de tal desempenho, reparou-se que as normalizações efetuadas concentraram os dados em torno do zero, em função do fato de existirem registros com várias ordens de grandeza superior ao de menor valor absoluto da série. Com isto a rede trabalhou com diferença de escala muito grande e com a maior parte dos registros próximos à zero. Como a rede tende a representar a maioria dos valores apresentados para treinamento, isto fez com que a generalização fosse prejudicada.

Em função do exposto, buscaram-se novos tipos de normalizações para os dados, visando uma maior dispersão dos pontos no intervalo desejado. Foram adotadas, então, as seguintes fórmulas:

- Normalização 1 = $\text{sqrt}(\text{abs}(x_i))$;
- Normalização 2 = $\exp(- (\text{abs}((x_i + \text{média}) / \text{variância})))$;
- Normalização 3 = $\exp(- ((x_i + \text{média}) ** 2 / \text{variância}))$;
- Normalização 4 = $\exp(- (\text{abs}((x_i + \text{média}) / \text{desvio padrão})))$.

Os gráficos referentes aos dados, limitados a 2 desvios padrões, após a aplicação da norma 1, estão mostrados a seguir, pois foram os que conseguiram a melhor dispersão e conjuntamente mais se aproximaram da curva normal (Gaussiana).

No caso das normas 2 e 3 as dispersões apresentadas não se mostraram adequadas pois além de se afastar da curva normal (Gaussiana) concentraram muitos pontos no limite superior e/ou inferior. Já a norma 4 apresentou uma dispersão quase que uniforme, inserindo, portanto, ruído na série, o que, obviamente, não é o desejado.

Em função dos motivos acima, escolheu-se a norma 1 como a melhor. Ainda assim, a título de testes, muitos dos experimentos realizados foram feitos utilizando-se das 4 normalizações.

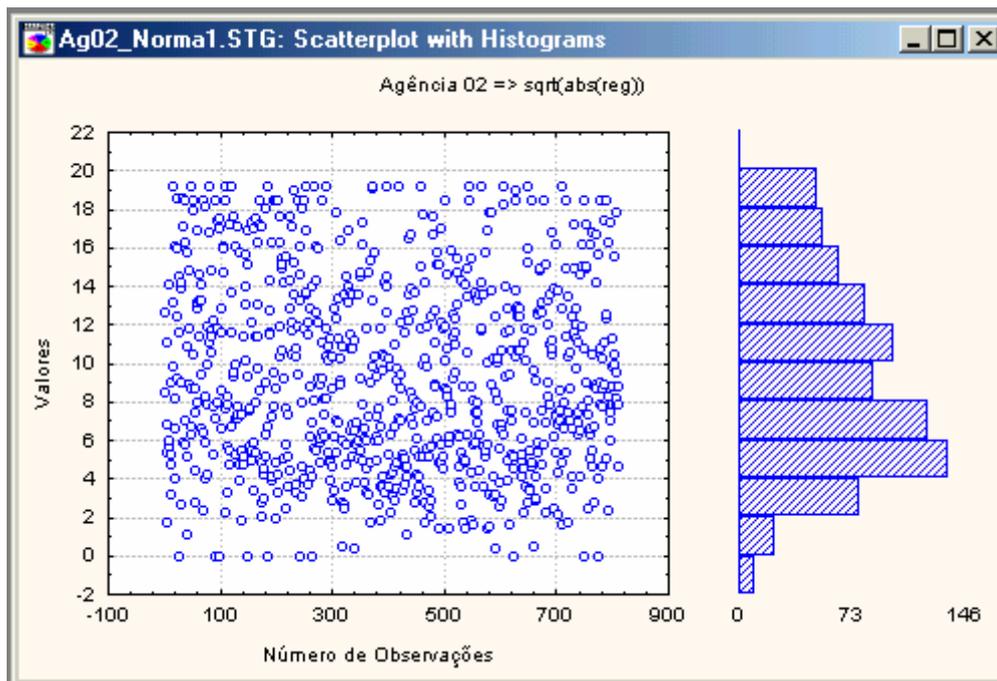


Gráfico 19: Agência 02 - $\sqrt{\text{abs}(\text{reg})}$.

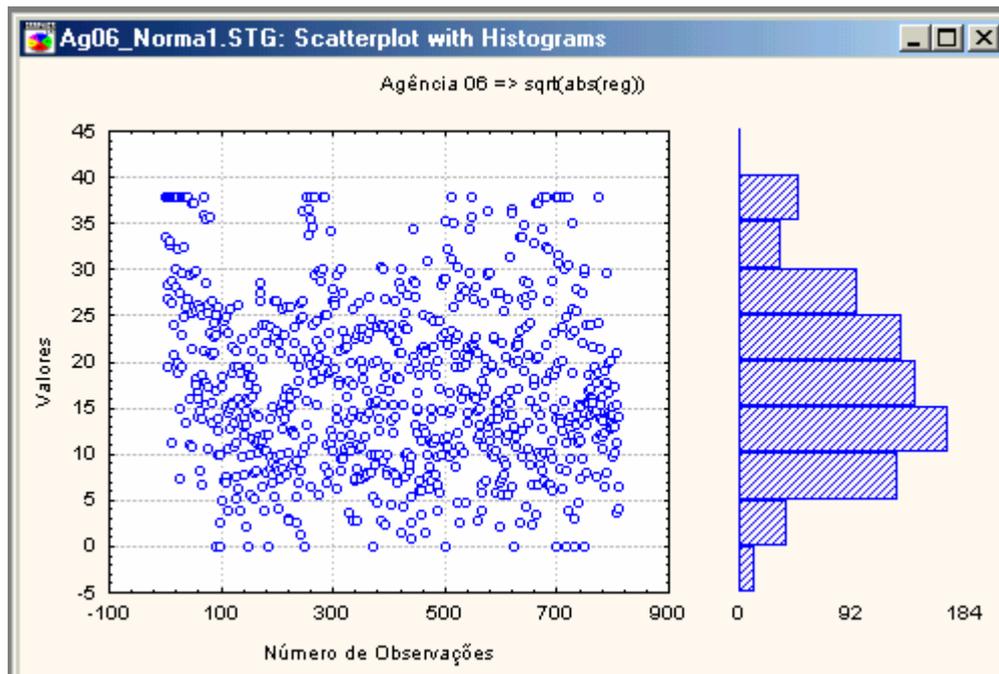


Gráfico 20: Agência 06 - $\sqrt{\text{abs}(\text{reg})}$

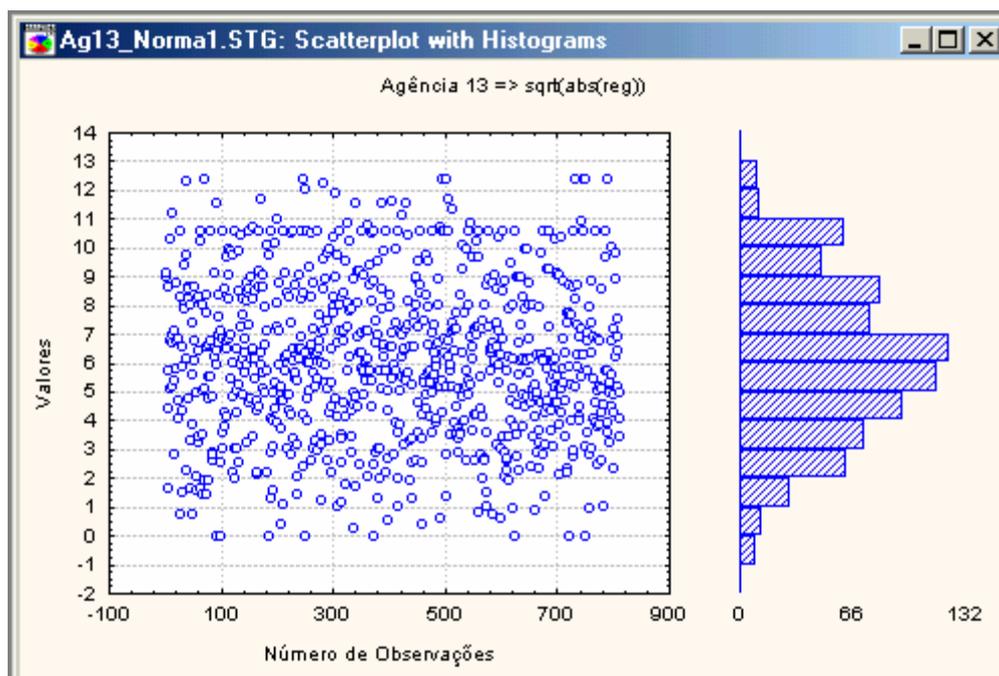


Gráfico 21: Agência 13 - $\sqrt{\text{abs}(\text{reg})}$.

O uso destas normalizações, inclusive a norma 2 que se mostrou mais adequada, não corresponderam à expectativa, ou seja, as performances das redes, após a realização de testes continuaram muito baixas, oscilando na ordem de 10% a 20% de acerto.

5. FERRAMENTAS DE MODELAGEM PARA PREVISÃO

Neste capítulo apresentam-se, nas seções 1 a 4, as técnicas clássicas de modelagem de séries temporais para, em seguida, nas seções 5 a 7, discorrer sobre as não clássicas (redes neurais artificiais e lógica nebulosa). Nas seções seguintes, 8 a 11, comenta-se sobre otimização, estrutura, treinamento e regras de aprendizado aplicadas às redes neurais artificiais. Por fim têm-se as conclusões.

5.1. FERRAMENTAS CLÁSSICAS

Uma vez que as condições econômicas e de negócios variam ao longo do tempo, os responsáveis pela condução da empresa precisam encontrar maneiras de se manterem a par dos efeitos que essas mudanças terão em suas operações. Uma técnica que pode ser empregada, como ajuda no planejamento de necessidades operacionais futuras, é a previsão. Existem diversas técnicas para se fazer previsão e todas elas se baseiam no pressuposto de que os fatores que influenciaram padrões da atividade no passado e no presente continuarão a fazê-lo no futuro (LEVINE et al., 2000, p. 630-631). Dentre as diversas técnicas existentes tratou-se, nesta dissertação, somente dos modelos de amortecimento exponencial (*exponential smoothing*) e ARIMA (*AutoRegressive Integrated Moving Average*) por serem bastantes difundidos na literatura e de amplo uso.

Define-se uma série temporal como qualquer conjunto de dados ordenados no tempo (STEVENSON, 1986, p. 412) (MORETTIN; TOLOI, 1987, p. 1). Seu estudo é de interesse de todas as áreas do conhecimento pois a variável tempo está presente em praticamente todas as séries, a exemplo de: temperatura média de determinada localidade, crescimento de bactérias, índice de fechamento da bolsa de valores, EEG (eletroencefalograma) etc.

Duas grandes classes de modelos são definidas no estudo de séries temporais. A primeira é de modelos univariados onde a série é prevista baseando-se somente nos seus valores passados; já a segunda classe, conhecida como modelos multivariados envolve, para a previsão dos valores da série, os seus próprios valores

passados como também outras variáveis. Neste último caso tem-se, como exemplo, a previsão do PIB (Produto Interno Bruto) brasileiro, aonde utiliza-se para cálculo não só os valores passados do PIB como também a taxa de inflação, cotação do dólar, desempenho da economia dos principais parceiros etc.

De uma maneira geral, o estudo de uma série temporal envolve dois aspectos: (i) analisar a série, ou seja, descrever seu comportamento e verificar as características existentes e relevantes e modelar seu mecanismo gerador, e (ii) prever acontecimentos futuros com base no modelo construído.

5.2. SÉRIES TEMPORAIS

Uma série temporal é composta, segundo o modelo clássico, por até 4 componentes (STEVENSON, 1986, p. 413): tendência, ciclos, sazonalidade e irregularidade. Não necessariamente todas as componentes descritas estão presentes em todas as séries. Pode-se encontrar séries com apenas 1, 2 ou 3 delas.

A tendência, como o próprio nome diz, está relacionada com o movimento dos dados, em longo prazo, para cima ou para baixo. São geralmente produzidas em função do crescimento/decrescimento constante da série. Por exemplo: o crescimento da população, queda constante do índice de inflação etc.

A segunda componente (variações cíclicas) e a terceira (sazonalidade) são padrões observáveis em períodos de tempo. A diferença básica está no fato de a primeira ser de longo prazo – períodos superiores a um ano, enquanto a segunda trata dos casos em intervalos de tempo iguais ou inferiores a um ano. Por exemplo: há estudos mostrando que o fenômeno meteorológico *El Niño* acontece em intervalos aproximados de 5 anos, ocorrendo, portanto, uma variação cíclica. Já a variação sazonal é relativa a períodos relativamente curtos, como o crescimento de vendas de casacos no inverno ou mesmo aparelhos de ar-condicionado nos meses de verão. Assim, a cada ano temos um pico de vendas nestes meses, caracterizando a sazonalidade.

Por último, encontramos as variações irregulares que decorrem de fatos anômalos não previsíveis, tais como catástrofes, guerras, greves etc. Enquadra-se neste item, também, tudo que não for passível de classificação nos padrões anteriores.

Na análise da série busca-se justamente decompô-la segundo os padrões descritos acima, para avaliação individual de cada componente. Procura-se investigar o

mecanismo gerador da série, isto é, descrever o seu comportamento, localizar periodicidades nos dados ou fazer previsões de valores futuros.

O estudo das séries temporais pode trazer informações importantes sobre o futuro, pois normalmente existe correlação entre as variáveis em diversos instantes. É claro que algum grau de incerteza virá agregado, pois o futuro nunca refletirá exatamente as ocorrências passadas mas, mesmo assim, a prática da previsão tem sido usada como auxílio ao planejamento e tomada de decisões.

5.3. AMORTECIMENTO EXPONENCIAL

Existem algumas variações do método de amortecimento exponencial, tais como: Simples, Holt, Winters e Brown, dentre outros. Todos estes métodos têm em comum o fato de o valor atual ser ponderado pela informação contida nos instantes anteriores. A existência de uma ou mais constantes de amortecimento irá determinar a taxa de decaimento da influência das observações passadas.

5.3.1. Amortecimento Exponencial Simples

A idéia geral do método de amortecimento exponencial é parecida com a de médias móveis, sendo que os pesos decrescem à medida que as observações estão mais longe no passado. A taxa de decréscimo dos pesos é determinada por uma ou mais constantes de amortecimento e a maior dificuldade na aplicação do método é justamente a escolha dessas constantes.

Na prática, os métodos de amortecimento são os mais usados em virtude da facilidade de implementação, à eficiência computacional e à sua razoável precisão (MONTGOMERY; JOHNSON; GARDINER, 1990, p. 81) (MORETTIN; TOLOI, 1987, p. 87).

Seja $Z = Z_1, Z_2, \dots, Z_n$ uma série temporal de tamanho n . Um modelo geral representativo no instante k genérico é dado por:

$$Z_k = \mu(k) + \varepsilon_k, \quad k = 1, 2, \dots, n \quad \text{Eq. (9)}$$

onde $\mu(k)$ representa o nível médio da série (função apenas do instante de tempo) e ε_k é um ruído aleatório com média zero e variância constante.

O cálculo da média móvel da série consiste em se obter a média aritmética das r observações mais recentes, onde r corresponde ao tamanho da “janela” adotada.

$$\hat{Z}_k = M_k = \frac{Z_k + Z_{k-1} + \dots + Z_{k-r+1}}{r} \quad \text{Eq. (10)}$$

Um dos fatores críticos nesta equação é justamente o de definir qual o tamanho ideal do parâmetro r a ser utilizado, ou seja, quantas observações devem ser computadas no cálculo da média móvel. Se $r = 1$, apenas a última observação é usada como previsão para todos os valores futuros. Por outro lado, se $r = n$ (tamanho da série) então todas as observações disponíveis são empregadas na estimação do nível médio da série e sua previsão será a média amostral da série.

Adota-se, geralmente, para escolha do parâmetro r , o inteiro positivo que minimiza a soma de quadrados dos erros de previsão um passo à frente, isto é, escolhe-se r de forma a minimizar:

$$S(r) = \min_r \left(\sum_{i=1}^n (Z_{k+i} - \hat{Z}_{k+i})^2 \right) \quad \text{Eq. (11)}$$

Assim, como pode ser visto na equação 10, M_k é uma estimativa do nível $\mu(k)$ que não pondera as ocorrências. Entretanto, é razoável imaginarmos que as observações mais recentes exercem maior influência nas previsões do que as mais antigas. Para tratar deste problema foi desenvolvido o método do amortecimento exponencial simples, onde o modelo proposto dá pesos maiores aos dados mais recentes.

Pode-se escrever o estimador do nível da série, produzido pelo método de médias móveis (equação 10), como:

$$\hat{Z}_k = M_k = \frac{Z_k + Z_{k-1} + \dots + Z_{k-r+1}}{r} \quad \text{Eq. (12)}$$

$$M_{k-1} = \frac{Z_{k-1} + Z_{k-2} + \dots + Z_{k-r}}{r} \quad \text{Eq. (13)}$$

$$M_k = \frac{Z_k}{r} + \frac{Z_{k-1} + Z_{k-2} + \dots + Z_{k-r+1}}{r} \quad \text{Eq. (14)}$$

$$M_{k-1} = \frac{Z_{k-1} + Z_{k-2} + \dots + Z_{k-r+1} + Z_{k-r}}{r} \quad \text{Eq. (15)}$$

Fazendo a diferença entre Eq. (6) e Eq. (7), tem-se:

$$M_k = M_{k-1} + \frac{Z_k - Z_{k-r}}{r} \quad \text{Eq. (16)}$$

onde M_{k-1} é a média móvel de tamanho r calculada no instante anterior. Verifica-se, portanto, que o estimador de nível \hat{Z}_k agrega uma informação atual, representada por Z_k (último valor da série).

O amortecimento exponencial simples é descrito matematicamente por:

$$\bar{Z} = \alpha Z_k + (1 - \alpha)\bar{Z}_{k-1}, \quad \bar{Z}_0 = Z_1 \quad k = 1 \dots n \quad \text{Eq. (17)}$$

ou

$$\bar{Z}_k = \alpha \sum_{j=0}^{k-1} (1 - \alpha)^j Z_{k-j} + (1 - \alpha)^k \bar{Z}_0 \quad k = 1 \dots n \quad \text{Eq. (18)}$$

onde \bar{Z}_k = valor exponencialmente amortecido

e α = constante de amortecimento, $0 \leq \alpha \leq 1$.

Considerando o termo $k+1$ da série equivalente à média dos k termos anteriores, temos

$$\hat{Z}_{k+1} = M_{k+1} + \frac{Z_k - M_{k+1}}{r} = M_{k+1} + \frac{Z_k}{r} - \frac{M_{k+1}}{r} \quad \text{Eq. (19)}$$

$$\hat{Z}_k = M_k = M_{k-1} + \frac{Z_k}{r} - \frac{M_{k-1}}{r} = \left(\frac{1}{r}\right)Z_k + \left(1 - \frac{1}{r}\right)M_{k-1} \quad \text{Eq. (20)}$$

Ou seja, com a reescrita do estimador vê-se que a nova fórmula tem a seguinte característica: (peso) * (informação atual) + (1 - peso) * (informação anterior).

Considerando que \hat{Z}_k é a previsão do nível da série efetuada no instante atual, podemos arrumar a equação anterior de forma a apresentá-la como:

$$\hat{Z}_k = \left(\frac{1}{r}\right)Z_k + \left(1 - \frac{1}{r}\right)M_{k-1} = \left(\frac{1}{r}\right)Z_k + \left(1 - \frac{1}{r}\right)\hat{Z}_{k-1} \quad \text{Eq. (21)}$$

onde o termo \hat{Z}_{k-1} nada mais é do que a média móvel de tamanho r calculada até a observação $k-1$, inclusive.

Assim, a previsão atual \hat{Z}_k é baseada na ponderação do valor mais recente da série (Z_k) e da última previsão para o nível (\hat{Z}_{k-1}). De uma forma geral, a equação 14 pode ser escrita como:

$$\hat{Z}_k = \alpha.Z_k + (1-\alpha).M_{k-1} = \alpha.Z_k + (1-\alpha).\hat{Z}_{k-1} \quad \text{Eq. (22)}$$

onde $\alpha \in [0,1]$ é a constante de amortecimento, que controla a taxa de decaimento da informação.

Se α é grande (próximo de 1) os pesos decaem rapidamente e uma observação vários passos anteriores não exercerá influência sobre a estimativa atual do nível. Ao contrário, com α pequeno uma observação vários passos anteriores ainda poderá ter influência na estimação do nível atual.

Segundo Morettin e Toloi (1987, p.99): *“Quanto menor for o valor de α mais estáveis serão as previsões finais, uma vez que a utilização de baixo valor de α implica que pesos maiores são dados às observações passadas e, conseqüentemente, qualquer flutuação aleatória, no presente, exercerá um peso menor no cálculo da previsão. Em geral, quanto mais aleatória for a série estudada, menores serão os valores da constante de alisamento.”*

Método de Holt

Também conhecido como “alisamento exponencial biparamétrico de Holt” é a técnica de amortecimento adequada para o tratamento de séries que apresentam tendência. Esse método é similar, em princípio, ao amortecimento exponencial simples. A diferença é que ao invés de amortecer só o nível, ele utiliza uma nova constante de amortecimento para “modelar” a tendência da série.

Uma série temporal não sazonal composta da soma de nível, tendência e resíduo aleatório, com média zero e variância constante pode ser representada matematicamente por:

$$Z_k = \mu(k) + \hat{a}_k + \varepsilon_k, \quad k = 1, 2, \dots, n \quad \text{Eq. (23)}$$

Os valores do nível e da tendência da série, no instante k , serão estimados por:

$$\hat{Z}_k = M_k = \alpha.Z_k + (1-\alpha)(M_{k-1} + \hat{a}_{k-1}) \quad 0 < \alpha < 1 \text{ e } k = 2, \dots, n \quad \text{Eq. (24)}$$

$$\hat{a}_k = \beta.(M_k - M_{k-1}) + (1-\beta)\hat{a}_{k-1} \quad 0 < \beta < 1 \text{ e } k = 2, \dots, n \quad \text{Eq. (25)}$$

onde α e β são as constantes de amortecimento a serem calculadas.

Como características importantes do método de Holt temos:

- As constantes de amortecimento usadas no cálculo do nível e tendência são diferentes;
- Na equação de atualização do nível, o novo nível é função de 3 variáveis: a observação mais recente (Z_k), a estimativa anterior do nível (M_{k-1}) e a taxa de tendência estimada no instante anterior (\hat{a}_{k-1});
- A taxa de tendência atualizada é função da taxa de tendência no instante anterior e também depende da diferença entre os níveis nos instantes $k-1$ e k .

Similar ao método de alisamento exponencial simples, a escolha dos valores das constantes é realizada através do cálculo da minimização da soma dos erros quadráticos de previsão para as duas variáveis envolvidas.

Método de Holt-Winters

O método de Holt-Winters, também conhecido simplesmente por método de Winters, visa estabelecer previsões para o caso de séries temporais que possuem um padrão de comportamento mais complexo, com sazonalidade e tendência.

Séries sazonais ocorrem com grande frequência no dia-a-dia. Safra agrícola, consumo de energia, temperatura média mensal, exportações e uma infinidade de outras séries são encontradas regularmente e o estudo das suas características é fundamental para gerar previsões consistentes.

As técnicas já apresentadas de amortecimento exponencial podem ser estendidas para modelar séries que apresentam tendência e sazonalidade, tratando-se as particularidades. Uma delas é o fato de que nos modelos anteriores o alisamento é realizado através da previsão de coeficientes polinomiais. No caso de séries com

sazonalidade esta abordagem tem que ser revista, pois dificilmente se consegue uma boa modelagem polinomial de série com sazonalidade ou com ciclo.

A principal dificuldade no uso do modelo de Winters é a determinação dos valores mais apropriados das constantes de alisamento. Aqui, como nos demais métodos, busca-se determinar essas constantes através da minimização da soma dos quadrados dos erros de ajustamento.

Existem dois tipos de modelos – o multiplicativo e o aditivo – e as suas utilizações são baseadas em três equações com constantes de alisamento diferentes, que são associadas às componentes de nível, tendência e sazonalidade.

Modelo Multiplicativo de Holt-Winters

Os modelos multiplicativos se caracterizam por terem a sazonalidade da série afetada pelo nível da mesma, ou seja, a série é uma resultante do produto das componentes individuais. Abaixo se tem um exemplo de série com sazonalidade multiplicativa onde se verifica, no gráfico à esquerda, o crescimento da curva e, à direita, a sua diminuição.

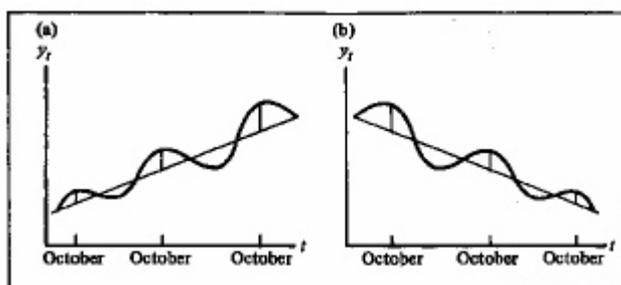


Gráfico 22: Modelo sazonal multiplicativo.

Seja uma série sazonal com período s . O modelo matemático que considera o fator sazonal ρ_k multiplicativo é:

$$Z_k = \mu(k) \cdot \hat{\rho}_k + \hat{a}_k + \varepsilon_t, \quad k = 1, 2, \dots, n \quad \text{Eq. (26)}$$

onde $\mu(k)$ representa o nível, \hat{a}_k a tendência, $\hat{\rho}_k$ a sazonalidade e ε_k o erro aleatório.

Precisa-se, portanto, estimar o nível (μ), a taxa de tendência (a) e os fatores sazonais ρ_k para todos os períodos k (meses, trimestres etc.) da série. A atualização dos parâmetros é feita sequencialmente e baseada nas seguintes equações de alisamento:

$$\hat{Z}_k = \alpha \left(\frac{Z_k}{\hat{\rho}_{k-1}} \right) + (1-\alpha) [\hat{Z}_{k-1} + \hat{a}_{k-1}] \quad 0 < \alpha < 1 \quad \text{Eq. (27)}$$

$$\hat{a}_k = \beta (\hat{Z}_k - \hat{Z}_{k-1}) + (1-\beta) \hat{a}_{k-1} \quad 0 < \beta < 1 \quad \text{Eq. (28)}$$

$$\hat{\rho}_k = \gamma \left[\frac{Z_k}{\hat{a}_k} \right] + (1-\gamma) \cdot \hat{\rho}_{k-1} \quad 0 < \gamma < 1 \quad \text{Eq. (29)}$$

A atualização dos fatores sazonais ($\hat{\rho}$) é feita apenas no período (mês, trimestre, etc) correspondente ao fator sazonal que se procura estimar.

Modelo Aditivo de Holt-Winters

Nos modelos aditivos o padrão sazonal não se altera à medida que o nível da série muda, conforme pode ser visto no gráfico abaixo.

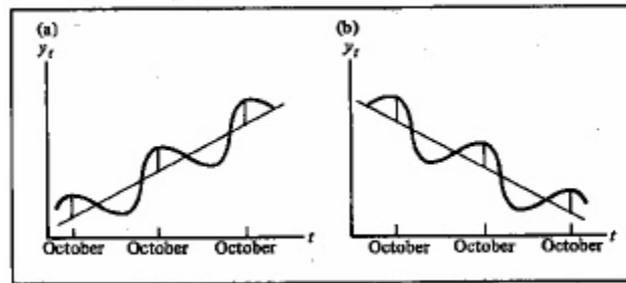


Gráfico 23: Modelo sazonal aditivo

A equação geral do modelo aditivo é dada por:

$$Z_k = \mu(k) + \hat{a}_k + \hat{\rho}_k + \varepsilon_k, \quad k = 1, 2, \dots, n \quad \text{Eq. (30)}$$

As equações de alisamento para o nível, tendência e sazonalidade do modelo aditivo são similares as anteriores, com algumas pequenas modificações, conforme se pode ver a seguir:

$$\hat{Z}_k = \alpha (Z_k - \hat{\rho}_{k-1}) + (1-\alpha) (\hat{Z}_{k-1} + \hat{a}_{k-1}) \quad 0 < \alpha < 1 \quad \text{Eq. (31)}$$

$$\hat{a}_k = \beta (\hat{Z}_k - \hat{Z}_{k-1}) + (1-\beta) \hat{a}_{k-1} \quad 0 < \beta < 1 \quad \text{Eq. (32)}$$

$$\hat{\rho} = \gamma [Z_k - \hat{Z}_k] + (1-\gamma) \cdot \hat{\rho}_{k-1} \quad 0 < \gamma < 1 \quad \text{Eq. (33)}$$

onde α , β e γ são as constantes de alisamento.

5.4. MODELO ARIMA - AUTOREGRESSIVE INTEGRATED MOVING AVERAGE

O modelo ARIMA, proposto por Box e Jenkins (BOX; JENKINS; REINSEL, 1994) faz parte da classe mais geral de modelos de previsão de séries temporais e podem ser aplicados em séries estacionárias ou não.

A sigla ARIMA em português significa auto-regressivo, integrado e médias móveis. Auto-regressivo corresponde à defasagem da série transformada, o termo médias móveis às defasagens dos erros aleatórios e o termo integrado ao processo de diferenciação da série original a fim de torná-la estacionária.

A condição de estacionariedade que é introduzida no modelo como forma simplificadora implica em considerar constantes a média e a variância da série temporal $\{Z_1, Z_2, \dots, Z_t\}$ e a covariância entre Z_t e Z_{t+k} dependente apenas do intervalo k .

A classe de modelos ARIMA é descrita matematicamente como (ABELÉM, 1994, p. 62)

$$Z_k = \phi_1 Z_{k-1} + \phi_2 Z_{k-2} + \dots + \phi_p Z_{k-p} + \varepsilon_k + \theta_1 \varepsilon_{k-1} + \theta_2 \varepsilon_{k-2} + \dots + \theta_q \varepsilon_{k-q} \quad \text{Eq. (34)}$$

Onde $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ são os q erros passados das séries. As constantes $\phi_1, \phi_2, \dots, \phi_p$ são os parâmetros auto-regressivos do modelo, enquanto $\theta_1, \theta_2, \dots, \theta_q$, são os parâmetros das médias móveis.

O modelo ARIMA possui 3 parâmetros necessários que são passados à série e representados como ARIMA(p,d,q) onde p é o número de termos auto-regressivos, d é o número de diferenças não sazonais e q é o número de termos médias móveis. Todos os parâmetros são inteiros e maiores ou iguais a zero.

É importante ressaltar que se $p = 0$ o modelo resultante será somente constituído de médias móveis, no caso de $d = 0$ temos que a série original já é estacionária, não necessitando de se realizar nenhuma diferenciação nos dados; por último, a hipótese de $q = 0$ significa que o modelo é puramente regressivo.

A construção de um modelo ARIMA se constitui de 3 etapas bem definidas (BOX; JENKINS; REINSEL, 1994, p. 181-182) (MONTGOMERY; JOHNSON; GARDINER, 1990, p.260):

- Identificação do modelo;
- Estimação; e
- Diagnóstico.

Alguns autores, a exemplo de Bowerman (BOWERMAN; O'CONNELL, 1979, p. 333) consideram que a construção do modelo é realizada em 4 passos – identificação, estimação, diagnóstico e previsão. Entretanto, a fase de previsão não pode ser considerada como integrante da construção do modelo pois o mesmo já foi desenvolvido nos passos anteriores. A previsão corresponde tão-somente à aplicação do modelo. Adotou-se, para efeito desta dissertação, somente os 3 passos mais aceitos como sendo os responsáveis pela fase de construção do modelo.

A identificação do modelo consiste na tentativa de se encontrar os parâmetros p , d e q que representem satisfatoriamente a série histórica em questão. Esta é a etapa considerada como sendo a mais crítica do método (MORETTIN; TOLOI, 1987, p. 219).

Inicialmente, a fim de se tornar a série estacionária, diferencia-se a série Z_k tantas vezes quanto necessárias. O número de diferenciações corresponde ao parâmetro d . Após esta etapa resta descobrir quais os melhores valores de p e q . Tais parâmetros são obtidos através dos cálculos das autocorrelações teóricas e autocorrelações parciais estimadas (correlações internas entre membros da série separados por intervalos de tempo constantes) cujos comportamentos devem imitar o das respectivas quantidades teóricas.

A segunda etapa de construção do modelo está relacionada com a estimação. Nesta fase já se tem identificado um modelo provisório para a série que foi obtida no passo 1. Calcula-se, então, os valores estimados dos coeficientes autoregressivos e/ou médias móveis e ajusta-se o modelo para a série em questão. Para realizar tal ajuste a maioria dos autores sugere o uso do método de mínimos quadrados. Entretanto, Morettin (MORETTIN; TOLOI, 1987, p. 237-238) trabalha com o cálculo de máxima verossimilhança para a realização de ajustes do modelo.

Finalmente, tem-se a fase de diagnóstico onde através do modelo construído nas etapas anteriores busca-se verificar se ele representa ou não corretamente os dados. A verificação é feita, na maioria das vezes, através da análise dos resíduos gerados pelo modelo construído.

Ao se ajustar um modelo ARIMA procura-se representar toda a estrutura de dependências dos dados existentes na série. Assim, a existência de resíduos no modelo não deve ter nenhuma relação com o tempo pois, caso exista, significa que o modelo construído não foi suficientemente abrangente. Nestes casos, deve-se retornar ao primeiro passo e iniciar novamente o processo de construção.

Após a conclusão com sucesso das etapas anteriores chega-se ao objetivo do trabalho que é o de realizar previsões k passos à frente baseado nos valores existentes na série sob estudo.

As previsões são obtidas diretamente a partir da equação do modelo construído nos passos anteriores. Seja $Z_k(j)$ a previsão de Z_{k+j} obtida no instante k . A equação de previsão é apenas a obtida da própria expressão do modelo para Z_t , escrita para o instante de interesse $t + j$.

5.5. FERRAMENTAS NÃO CLÁSSICAS

As técnicas da estatística clássica para predição e tomada de decisão não apresentaram bons resultados para muitas aplicações que possuem tamanho de amostras restrito e com alguma não linearidade no conjunto de dados e, em função disto, começaram a surgir alguns estudos na área de inteligência computacional, tais como redes neurais artificiais e lógica nebulosa.

A abordagem com redes neurais ganhou especial atenção em vista da sua habilidade de aprendizado e capacidade de generalização, associação e busca paralela. Estas qualidades as tornam capazes de identificar e assimilar automaticamente as características mais marcantes das séries, tais como sazonalidade, periodicidade, tendências, entre outras (ABELÉM, 1994, p. 26-27).

A lógica nebulosa é uma forma elegante e eficaz para solucionar problemas, pois sistemas construídos com base neste paradigma são mais compactos e baratos que aqueles que utilizam soluções convencionais. Além disso, esse paradigma possui a capacidade de processar dados inexatos ou parcialmente corretos. De fato, os sistemas nebulosos vêm sendo utilizados em diversos tipos de aplicações comerciais, inclusive na área de previsão de séries temporais.

5.6. REDES NEURAIS ARTIFICIAIS

A permanente tentativa de o homem tentar reproduzir a natureza em todos os seus aspectos levou a, entre outras descobertas, modelar matematicamente o funcionamento de um neurônio biológico. Dessa forma, a partir do século XIX, com o trabalho de identificação anatômica e mapeamento do neurônio feito pelo espanhol Tamón y Cajal (KOVÁCS, 1996, p. 13), nasceu a teoria de Redes Neurais.

Em 1943 Warren McCulloch e Walter Pitts publicaram o artigo "*A Logical Calculus of the Ideas Immanent in Nervous Activity*" que é considerado o trabalho pioneiro tratando do assunto (BRAGA; CARVALHO; LUDEMIR, 1998, p. 2). Embora rudimentar se comparado aos modelos existentes atualmente, foi inovador e abriu caminho para este novo ramo da ciência. Nele, McCulloch e Pitts enfocam a capacidade computacional do neurônio biológico.

Em fins da década de 50, Frank Roseblatt desenvolveu, na Universidade de Cornell, o modelo de rede neural conhecido como Perceptron. É uma rede de múltiplos neurônios do tipo discriminadores lineares. A topologia original era formada por uma camada de entrada, um nível intermediário composto pelas unidades de associação e um nível de saída formado pelas unidades de respostas. Embora possua três níveis, esta topologia é conhecida como sendo de uma única camada, pois somente o nível de saída possui propriedades adaptativas.

Na mesma época em que Roseblatt realizava suas pesquisas com o Perceptron, o pesquisador Bernard Widrow, da Universidade de Stanford, desenvolveu um outro modelo neural linear chamado ADALINE (*ADaptive LINEar Element*), posteriormente popularizado como *ADaptive Linear NEuron*. Trata-se de um modelo conceitualmente simples mas que apresentou um princípio de treinamento extremamente poderoso, conhecido como regra Delta, que deu origem ao algoritmo de *back-propagation*. Mais tarde tal conceito foi generalizado e atualmente, aplica-se esta regra em modelos neurais mais complexos. Do perceptron de Roseblatt derivou-se as redes do tipo MLP (*Multilayer Perceptron*) que, juntamente com o algoritmo de aprendizagem *back-propagation*, é hoje um dos modelos mais utilizados.

Até poucas décadas atrás ainda se buscava a total similaridade entre os modelos neurais e o cérebro humano. Dessa forma, as definições usadas para as redes

neurais sempre procuravam manter uma estreita relação com o cérebro. Nessa linha, Teuvo Kohonen apresenta as redes neurais como (KOHONEN, 1987, I-81).

“The artificial neural networks are massively parallel interconnected networks of simple (usually adaptive) elements and their hierarchical organizations which are intended to interact with the objects of the real world in the same way as the biological nervous systems do”.

Entretanto, atualmente os modelos neurais são tão grosseiros em relação ao conhecimento existente do comportamento do cérebro que dificilmente pode-se justificar a palavra neural. Hoje em dia o termo neural é mais usado por razões históricas, pois é oriundo de trabalhos realizados por pesquisadores das áreas de biologia e psicologia. (NASCIMENTO; YONEYAMA, 2002, p. 112).

Uma definição mais atual para as redes neurais artificiais pode ser encontrada no artigo apresentado no XV Congresso de Pesquisa e Ensino em Transportes, por Oliva et al. (OLIVA; CYBIS; PRETTO, 2001, p. 284):

“Uma rede neural artificial é uma estrutura formada por unidades de processamento comumente chamadas de nós, neurônios ou células, interconectadas por arcos direcionais, também chamados de ligações, conexões ou sinapses. De um modo geral, cada neurônio artificial recebe um conjunto de entradas ou sinais x_i , calcula uma média ponderada entre eles, n_k , e utiliza alguma função, $j(n_k)$, que determinará sua ativação ou não, para produzir uma saída, y_k , que será a entrada para os outros neurônios. Seguindo o mesmo processo em toda a rede, fica caracterizado o processamento paralelo distribuído que se constitui em um dos maiores benefícios do paradigma neural, ou seja, a não-linearidade difundida por toda a rede”.

Atualmente, as redes neurais são aplicadas, em sua maioria, nos problemas envolvendo clusterização, previsão e reconhecimento. Trataremos aqui somente os modelos de redes cujo objetivo é o de previsão pois as demais fogem ao âmbito deste trabalho. Estas redes são conhecidas, também, como aproximadores de funções e, neste universo, encontram-se as redes MLP (*Multilayer Perceptron*) e “Elman”, entre outras.

5.6.1. O Modelo Multilayer Perceptron

Uma rede composta de uma única camada de perceptrons resolve somente problemas que são linearmente separáveis. Portanto o escopo de sua utilização torna-se muito restrita para implementação na maioria das situações apresentadas.

A evolução natural trouxe as topologias de várias camadas, conhecidas como MLP (*Multilayer Perceptron*). Junto com esta nova arquitetura vieram problemas que necessitaram de soluções a fim de viabilizá-la e o primeiro grande desafio a ser superado foi o de como treinar uma rede de várias camadas. Poderia-se simplesmente dividir a rede em sub-conjuntos e treiná-los separadamente. Experiências neste sentido mostraram que este corte não é possível ou a complexidade no tratamento da rede cresce tanto que a inviabiliza (BRAGA; CARVALHO; LUDEMIR, 1998, p. 49).

Passou-se, então, a pensar em como treinar a rede por inteiro. Para tanto, há a necessidade de se mensurar o erro na camada interna e, de preferência, que estas camadas sejam adaptativas, ou seja, que os pesos de suas conexões possam ser devidamente atualizados. Para solucionar este problema, Rumelhart et al. propuseram o algoritmo de treinamento conhecido como *back-propagation* (RUMELHART; HINTON; WILLIAMS, 1986, p. 533-536). Apesar da eficiência deste método, ele se restringe aos casos em que a função é contínua, diferenciável e, de preferência, não decrescente (HAYKIN, 1999, p. 168).

O desenvolvimento das RNAs foi baseado no estudo de como o cérebro humano processava as informações, resultando na criação de modelos que procuravam simular tal processamento — feito pelos neurônios — por meio do uso de Elementos de Processamento (EP). Assim, a RNA é uma estrutura de processamento de dados em paralelo e distribuído, formada pela união entre si de vários EPs, através de conexões unidirecionais. O EP pode possuir R entradas porém só terá uma saída que poderá ser conectada a outros EPs. Todo o processamento que ocorre em cada EP é local, dependendo apenas dos valores dos sinais de entrada em suas conexões e dos valores armazenados na memória local do elemento.

A figura adiante mostra um modelo genérico do neurônio artificial como é aceito atualmente. Neste modelo, P_1 a P_R são os sinais de entrada e W_{i1} a W_{iR} são os pesos das R entradas (ou R conexões vindas de outros neurônios), responsáveis pela ponderação dos sinais de entrada, $a(.)$ é a função de ativação, NET_i é o estado de

ativação do i -ésimo EP, $p(\cdot)$ é a função de propagação, $F(\cdot)$ a função de transferência, O_i a saída do i -ésimo EP podendo ser transmitida para outros m neurônios através de m conexões. A interconexão de diversos EPs como este forma uma arquitetura de RNA.

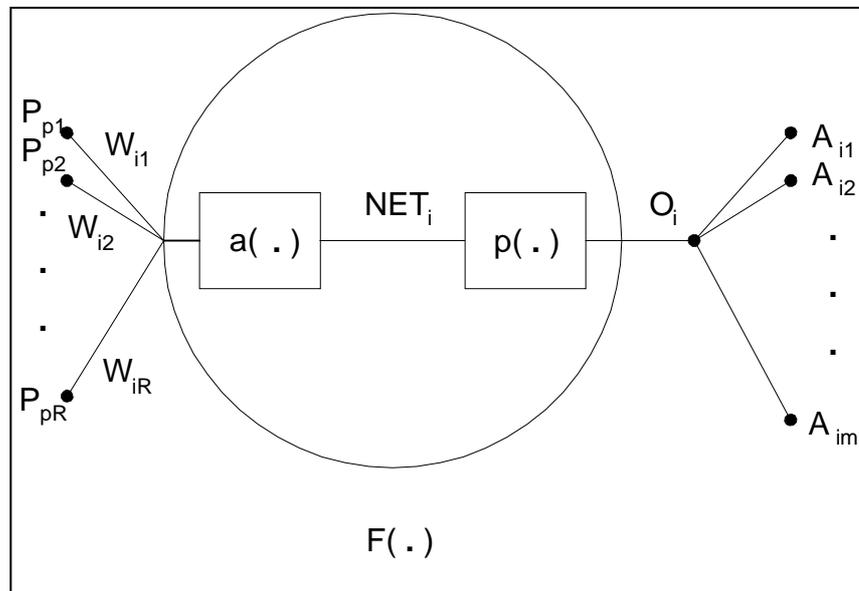


Figura 2: Modelo genérico do neurônio artificial

5.6.2. Treinamento e Operação de Redes Neurais do Tipo MLP (MultiLayer Perceptron)

O primeiro modelo de um neurônio artificial, proposto por McCulloch e Pitts em 1943, era uma simplificação do neurônio biológico e está mostrado adiante.

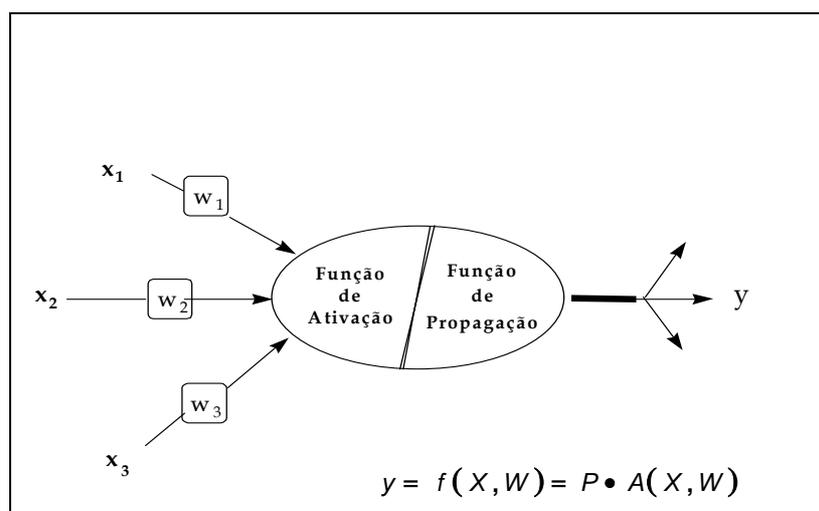


Figura 3: Modelo do Neurônio

A representação matemática proposta consiste em uma função de ativação composta pelo produto dos pesos com as entradas e uma função de propagação (*threshold*) que excita ou inibe o sinal.

A proposta de McCulloch era de um dispositivo binário onde poderia existir uma saída ou não e as entradas seriam excitatórias ou inibitórias. O modelo calculava a soma ponderada das entradas com o seu ganho (peso- W), positivos nos casos excitatórios e negativos no contrário. Se o resultado fosse igual ou maior que o *threshold*, então haveria uma saída, do contrário não se teria o pulso.

A partir deste modelo vários outros foram desenvolvidos utilizando-se outras funções de ativação diferentes daquela proposta por McCulloch e Pitts (função de limiar). A saída do neurônio também deixou de ser simplesmente 0 ou 1, aceitando-se qualquer valor.

Em 1949 o biólogo D.O.Hebb propôs que a intensidade das conexões sinápticas deveria ser alterada apenas em função dos erros detectáveis localmente. Por esta regra, a alteração do i -ésimo elemento fica dependente tão-somente do produto da i -ésima entrada pela saída do neurônio pré-sináptico. Em outras palavras, Hebb propôs que o peso de uma conexão sináptica só deve ser ajustado se houver sincronismo entre eles, ou seja, se os neurônios forem ativados sincronamente teremos um fortalecimento da sinapse, do contrário teremos um enfraquecimento ou mesmo a eliminação desta sinapse.

Em termos matemáticos, a proposta de Hebb pode ser expressa através da seguinte equação:

$$\Delta w_{ij} = \lambda \cdot O_i \cdot X_j, \quad \text{para } j = 1 \quad \text{Eq. (35)}$$

onde λ representa a taxa de aprendizado e O_i a saída do neurônio pré-sináptico

Posteriormente, Rosenblatt usou o princípio hebbiano para alterar os parâmetros de treinamento do peso (W) de um discriminador linear. Este modelo ficou conhecido como perceptron.

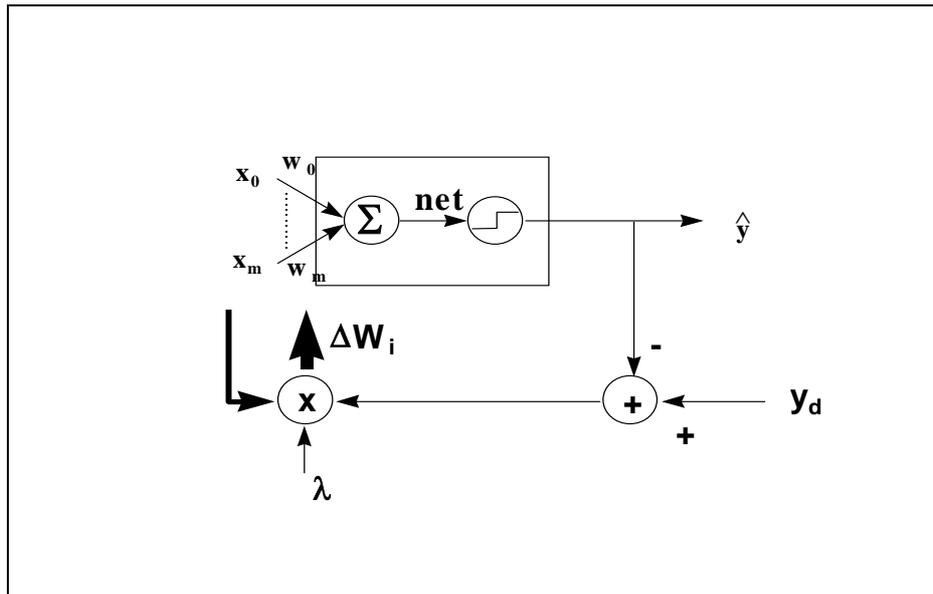


Figura 4: Representação esquemática do Perceptron

Nota-se, abaixo, que a alteração proposta por Rosenblatt resume-se à troca do parâmetro O_i da equação de Hebb, pelo erro verificado entre o padrão desejado y_d e o estimado \hat{y} .

$$\Delta w_i = \lambda.(y_d - \hat{y}).X, \quad \text{onde } \lambda = \text{taxa de aprendizagem} \quad \text{Eq. (36)}$$

Na mesma época em que Rosenblatt divulgava seu modelo neural, Bernard Widrow apresentava o modelo Adaline (figura abaixo). Ambos os modelos são baseados na mesma idéia, ou seja, ter-se elementos de processamento executando operações de soma ponderada e posterior comparação com um valor de limiar.

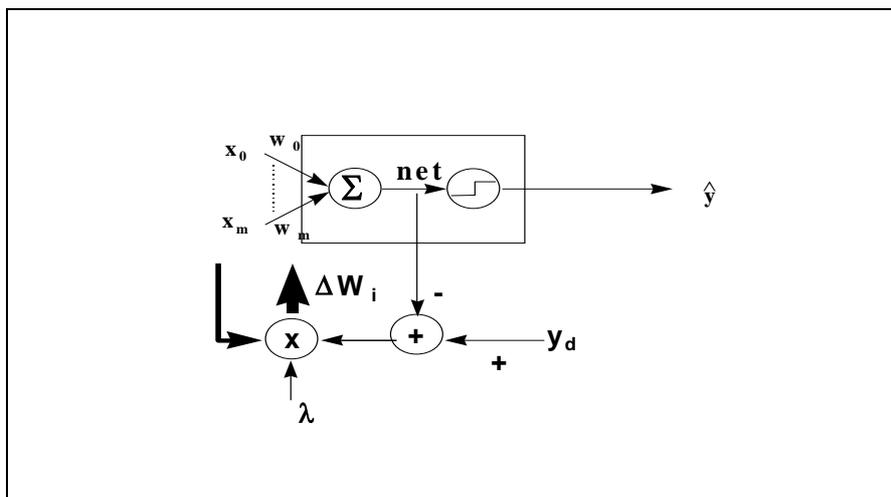


Figura 5: Modelo do Adaline

$$\Delta w_i = \lambda.(y_d - \hat{y}).X, \text{ onde } \lambda = \text{taxa de aprendizado} \quad \text{Eq. (37)}$$

O algoritmo de treinamento de uma RNA do tipo MLP mais utilizado atualmente é conhecido como *back-propagation* e foi o responsável pelo ressurgimento das pesquisas em redes neurais, pois a partir da sua concepção o maior problema do perceptron, que era a impossibilidade de se treinar neurônios em camadas intermediárias, foi solucionado. A grande novidade introduzida neste algoritmo foi a substituição da função degrau utilizada no perceptron original por uma função derivável, no caso a sigmóide. Com esta alteração a regra de aprendizado pode ser estendida para as demais camadas.

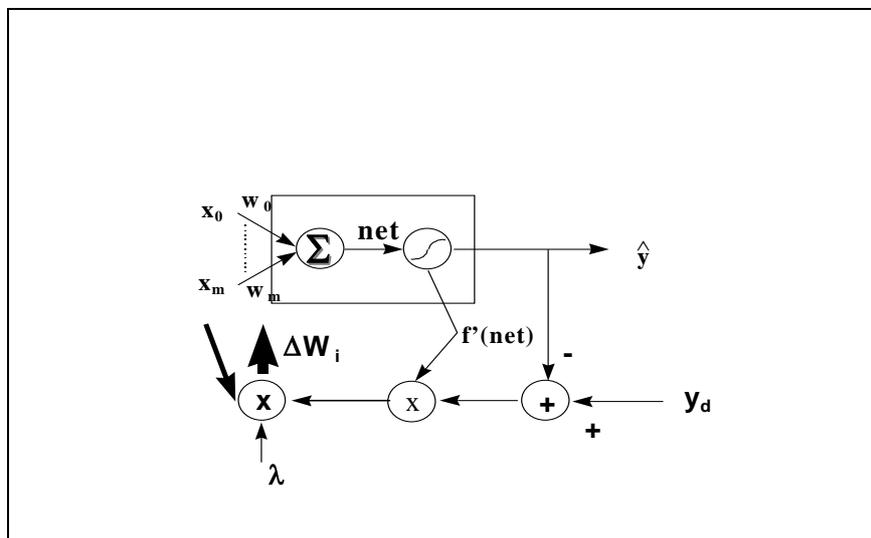


Figura 6: Modelo de neurônio utilizando o *back-propagation*

$$\Delta w_i = \lambda.(y_d - \hat{y}).f'(net).X \quad \text{Eq. (38)}$$

Apesar do algoritmo de *back-propagation* ser eficiente do ponto de vista da correção dos erros dos neurônios nas camadas escondidas, a sua convergência ocorre de forma muito lenta. Ele atua nos dois sentidos: da entrada para a saída, conhecido como fase *forward* e, posteriormente, da saída para a entrada, chamado de *backward*.

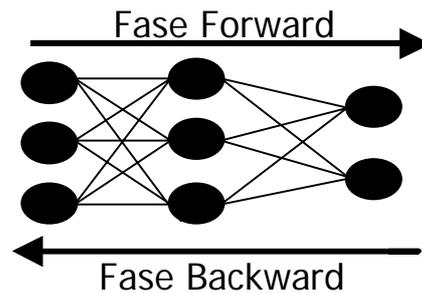


Figura 7: Fluxo de processamento do back-propagation

A fase *forward* define o erro gerado na saída da rede. De posse deste valor inicia-se a fase *backward* que, no sentido inverso do fluxo vai atualizando os pesos, baseados no erro verificado na saída da rede.

O fluxo de treinamento, então, ocorre da seguinte forma:

Fase *forward*:

- Apresenta-se a entrada para a rede;
- Cada neurônio calcula seu sinal de propagação e o transmite para o seguinte;
- Na saída compara-se o valor obtido com o desejado e calcula-se o erro produzido.

Fase *backward*:

- Os neurônios da camada de saída ajustam seus pesos visando minimizar o erro;
- O sinal é transmitido para os neurônios da camada imediatamente anterior que, por sua vez, também corrigem seus respectivos pesos;
- Este procedimento se repete até atingir a camada de entrada. A seguir novo elemento é apresentado para a rede e o procedimento de *forward* é reiniciado.

5.6.3. O Modelo “Elman”

Fenômenos temporais são encontrados em atividades diárias dos seres humanos. Dessa forma, o tempo é um importante parâmetro a ser mapeado nos modelos de redes a fim de se poder realizar várias tarefas, tais como: reconhecimento de fonemas, processamento de sinais etc.

A computação convencional assim como a maioria dos algoritmos de treinamento de redes neurais não estão aptos a tratar automaticamente da temporalidade dos dados. As redes recorrentes, ou redes com laços (*loops*), são uma tentativa para remediar essa situação.

Redes do tipo recorrentes foram descritas por Jordan (1988) e Elman (1990) e possuem unidades de contexto, que são elementos de processamento que memorizam atividades passadas, simulando um certo tipo de memória de curto-prazo. Estas unidades de contexto são importantes quando o processo de aprendizagem dos exemplos ocorre ao longo do tempo, ou seja, sempre que os valores passados da rede influenciam os valores presentes.

Nas redes “Elman”, além das unidades de entrada, intermediárias e de saída, há também unidades de contexto. As unidades de entrada e saída interagem com o ambiente externo, enquanto as unidades intermediárias e de contexto não o fazem. As unidades intermediárias podem ter funções de ativação lineares ou não lineares e as unidades de contexto são usadas apenas para memorizar as ativações anteriores das unidades intermediárias e podem ser consideradas como atrasos no tempo (BRAGA; CARVALHO; LUDEMIR, 1998, p. 202).

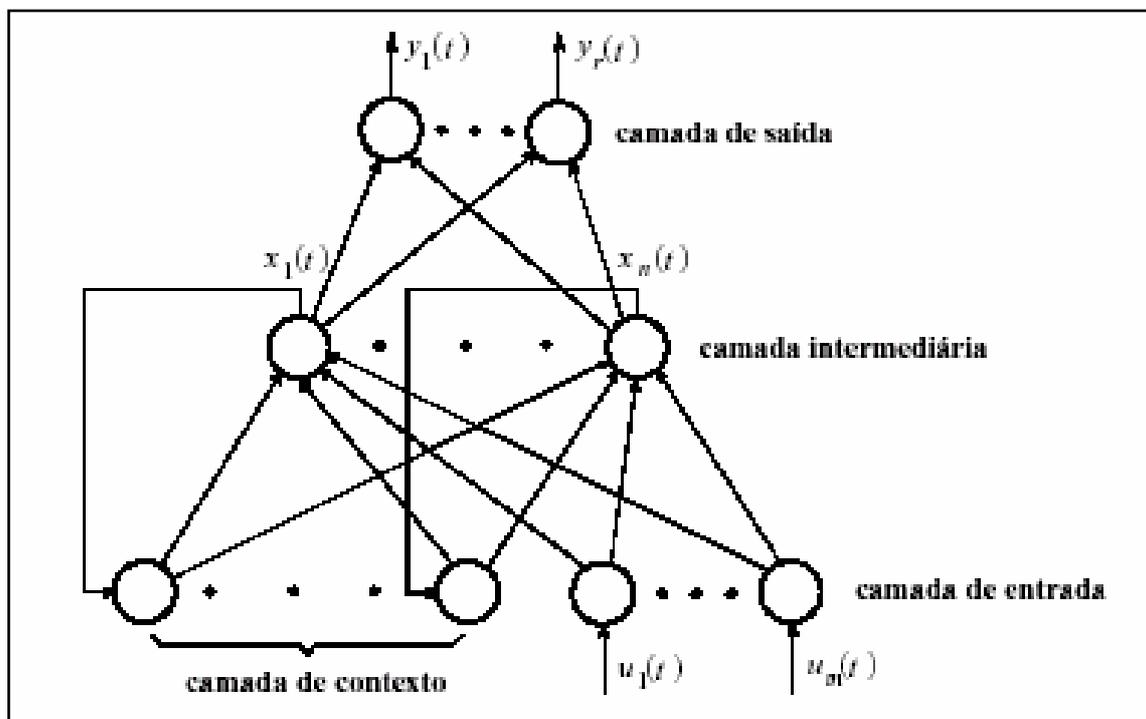


Figura 8: Modelo de uma RNA do tipo Elman

O esquema da rede recorrente mostrado na figura acima é similar ao de uma rede de retropropagação. A principal diferença reside na adição da camada de recorrência ou de contexto, que faz com que a saída da camada escondida seja uma nova entrada dessa mesma camada, representando um intervalo de tempo anterior.

As redes recorrentes podem ter uma ou mais camadas, mas a sua particularidade reside no fato de possuir conexões que partem da saída de uma unidade em direção a uma outra unidade da mesma camada ou de uma camada anterior a esta. Este tipo de conexão permite a criação de modelos que levam em consideração aspectos temporais e comportamentos dinâmicos, onde a saída de uma unidade depende de seu estado em um tempo anterior.

5.6.4. *Redes com Atraso no Tempo*

O tempo constitui um elemento essencial no processo de aprendizagem. Ele pode ser contínuo ou discreto. Independentemente da sua forma, o tempo é uma entidade ordenada que é básica para muitas tarefas encontradas na prática, como mercado de ações, processamento de sinais, evolução das vendas de um produto etc. É através da incorporação do tempo na operação de uma rede neural que ela é capacitada a seguir as variações estatísticas em processos não-estacionários como flutuações em preços do mercado de *commodities*, consumo de energia elétrica e fluxo de caixa de uma empresa, dentre outros.

Muitos algoritmos de treinamento das redes neurais artificiais não são capazes de implementar mapeamentos dinâmicos como, por exemplo, o algoritmo *backpropagation*, que pode apenas aprender mapeamentos estáticos. Para que uma rede neural seja dinâmica, ela deve ter “memória”. A memória pode ser de curto ou de longo prazo, dependendo do tempo de retenção. A memória de longo prazo é inserida em uma rede neural através de aprendizagem supervisionada, pela qual o conteúdo de informação do conjunto de dados de treinamento é armazenado (parcialmente ou totalmente) nos pesos sinápticos da rede. Entretanto, se a tarefa considerada tiver uma dimensão temporal, necessita-se de alguma forma de memória de curto prazo para tornar a rede dinâmica. Uma forma simples de inserir memória de curto prazo na estrutura de uma rede neural é através de atrasos de tempo, que podem ser implementados a nível sináptico dentro da rede ou na camada de entrada da rede

(HAYKIN, 2001, p. 686). Um artifício utilizado para o processamento temporal através destas redes envolve o uso de janelas de tempo, onde a entrada da rede utiliza trechos dos dados temporais como se eles formassem padrões estatísticos (BRAGA; CARVALHO; LUDEMIR, 1998, p.198).

Esta dissertação utilizou, em todos os experimentos envolvendo redes do tipo MLP, uma janela de tempo, visando simular um processamento dinâmico. Foram testadas várias janelas a fim de se verificar o tamanho ideal para a entrada de dados e a janela que melhor se adequou ao problema foi a com 5 registros, apesar de não haver fornecido resultados satisfatórios.

Basicamente, trabalhou-se, com rede Elman, que é recorrente em sua própria arquitetura e a MLP dinâmica. Cabe comentar que outros modelos (SONN, Radial Basis Function-RBF, Kohonen) foram testados mas, em decorrência do fraco desempenho, não se deu prosseguimento.

5.7. ESTRATÉGIA, TREINAMENTO E CONSTRUÇÃO DE REDES NEURAIS

As redes neurais do tipo *MLP* são extremamente eficientes nos problemas de classificação, previsão e outros mas, freqüentemente, requerem muito tempo no seu treinamento. Visando reduzir o esforço computacional, bem como melhorar as chances de convergência dos resultados, pesquisadores vêm estudando e avaliando estratégias e métodos de treinamento, como forma de melhorar a performance do algoritmo de treinamento.

Vários parâmetros influenciam decisivamente na performance da rede, tais como a quantidade de neurônios, função de transferência, o esforço computacional despendido, o tempo de processamento, taxa de aprendizado, algoritmo de treinamento, taxa de momento etc.

Em razão da grande variedade de parâmetros a serem ajustados, as redes neurais não possuem uma forma única, ou seja, para um mesmo problema pode-se ter várias soluções similares. Além disso, melhorias nos resultados geralmente podem ser obtidos mediante otimizações no modelo, mudanças nas configurações, alterações nos valores dos parâmetros etc.

A construção de um modelo neural para atender satisfatoriamente um problema específico é uma tarefa não trivial, apesar de este modelo teoricamente

sempre existir, conforme afirma Rumelhart et al em seu artigo de 1994 (RUMELHART; WIDROW; LEHR, 1994, p. 90).

“If we have enough connections from the input units to a large enough set of hidden units, we can always find a representation that will perform any mapping from input to output through these hidden units”.

A determinação da arquitetura e dos diversos parâmetros envolvidos são decisões que devem ser tomadas e avaliadas criteriosamente com o intuito de se obter a melhor relação possível entre a eficiência e a eficácia do modelo. Portanto, caberá ao responsável pela modelagem a tarefa de encontrar os parâmetros corretos para resolução do problema em estudo.

Segundo Maren (MAREN et al., 1990, p. 234) a melhoria da performance das redes passa pela avaliação de conceitos inseridos em 3 grandes áreas:

- estrutura da rede (arquitetura);
- dinâmica ou treinamento da rede;
- regras de aprendizado.

No primeiro grupo têm-se a escolha do número de camadas, da quantidade de neurônios e de conexões entre os nós e definição das funções de ativação e propagação. Segundo os autores, modificações implementadas nas funções de transferência podem melhorar significativamente a performance da rede e cita como exemplo o estudo de Stornetta (STORNETTA; HUBERMAN, 1990) através do qual com um simples deslocamento de $\frac{1}{2}$ na função logística conseguiu redução no tempo de treinamento entre 30% e 50%. Tal fato aconteceu em função da simetria em torno do 0 (zero) que foi obtida a partir do deslocamento dos valores.

Outra grande área de estudo a ser avaliada é referente à dinâmica ou treinamento da rede. Nela os autores englobam o tratamento que deve ser dispensado na representação dos registros de entrada e técnicas de pré-processamento.

Por último, tem-se a área relativa às regras de aprendizado onde insere-se os trabalhos visando os algoritmos de treinamento. Aqui têm-se dois subitens: o primeiro é relativo aos parâmetros ajustáveis do algoritmo, tais como o momento para o caso do *back-propagation* e a taxa de aprendizado. O segundo está relacionado com o uso de algoritmos para aumento da eficiência do treinamento, a exemplo do *quickprop* (FAHLMAN, 1988).

Em maior ou menor grau outros autores também avaliam como crítica a determinação dos parâmetros da rede para o sucesso do modelo. Zhang (ZHANG; PATUWO; HU, 1988, p. 42) considera como fatores importantes a serem observados a definição do número de camadas, número de neurônios em cada camada, número de conexões entre os nodos bem como a definição das funções de ativação em cada camada. O algoritmo apropriado de treinamento, transformação de dados e métodos de normalização também são citados como itens passíveis de análise criteriosa. Por fim, mas não menos importante, têm-se a definição das medidas de performance e escolha dos conjuntos de treinamento, validação e teste.

5.8. ESCOLHA DA ESTRUTURA DA REDE (ARQUITETURA)

Após a definição da arquitetura inicial a ser utilizada, através de sucessivos refinamentos do modelo, procura-se chegar a um dimensionamento da rede que seja o menor possível sem, no entanto, se perder de vista o objetivo principal: a boa capacidade de generalização do modelo aliada a melhor aproximação possível dos resultados estimados com os reais.

Haykin (HAYKIN, 1999, p. 218) sugere duas abordagens para se alcançar a melhor estrutura:

- Através do crescimento da rede: neste caso começa-se com uma quantidade pequena de neurônios e vai-se adicionando novos neurônios e novas camadas ao modelo. A cada nova inserção avalia-se a performance para verificar se está melhorando ou não;
- Através da poda da rede: este método é exatamente o inverso do anterior, ou seja, inicia-se com uma rede grande e vai-se retirando neurônios e avaliando a performance.

As técnicas acima, apesar de simples, requerem alguns cuidados na sua aplicação. No primeiro caso o crescimento acentuado da rede pode levá-la a memorizar os padrões de treinamento, ao invés de extrair as características gerais que permitirão a generalização ou reconhecimento de padrões não vistos durante o treinamento. Tal fato, conhecido como *overfitting* pode ser evitado estimando-se o erro de generalização durante o processo de aprendizado da rede, ou seja, divide-se o conjunto de dados em dois grupos: treinamento e validação e durante a fase de treinamento da rede vai-se

medindo o erro do conjunto de validação. Quando este começa a subir deve-se parar o procedimento (BRAGA; CARVALHO; LUDEMIR, 1998, p.56).

No caso inverso, ou seja, na poda, o procedimento adotado corresponde a ir-se eliminando neurônios e testando a performance da rede. A supressão em demasia de elementos neuronais pode forçar a rede a gastar tempo em excesso tentando encontrar uma representação ótima, pois as unidades utilizadas podem ficar sobrecarregadas, tendo que lidar com um elevado número de restrições. Este problema, conhecido como *underfitting*, caracteriza-se pelo fato de a rede não conseguir apresentar uma performance adequada em função do baixo grau de liberdade do modelo. A solução para se evitar este problema passa, também, pelo acompanhamento do cálculo do erro no conjunto de validação que, quando começa a aumentar, indica que está na hora de se parar o procedimento.

A seguir tem-se a representação de três curvas referentes a treinamentos com bases de dados distintas e hipotéticas onde, na primeira à esquerda, o treinamento ocasionou um *underfitting*. Vê-se, neste caso, que o mapeamento da função original (função contínua no gráfico) realizado pela rede gerou pontos (representados pelo “X”) muito afastados da curva real, tendo como consequência um erro médio elevado.

O segundo caso corresponde a um treinamento eficaz, onde a maioria dos pontos mapeados pela rede se encontra sobre a curva com alguns outros ligeiramente afastados.

Por último, tem-se o gráfico referente ao *overfitting* em que todos os pontos estão sobre a curva original, caracterizando uma memorização. A generalização, neste caso, é muito pobre.

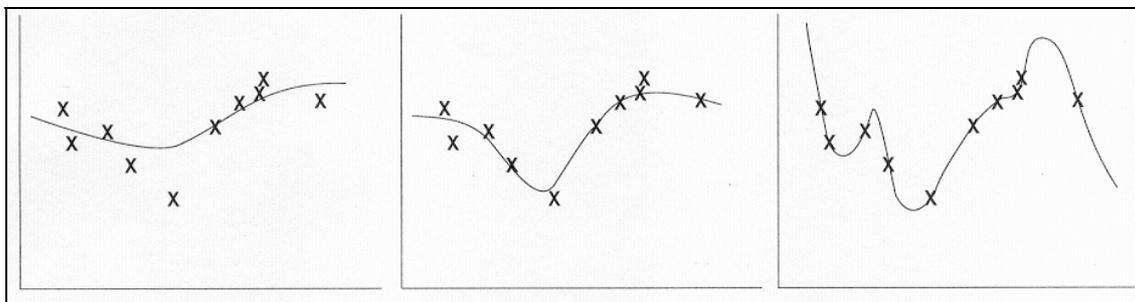


Figura 9: Representação do resultado de diversos treinamentos: à esquerda tem-se o *underfitting*, ao centro um treinamento correto e à direita o *overfitting*.

Considerando o erro médio como parâmetro de avaliação da performance do treinamento, vê-se que ele é alto para o caso do *underfitting*. A solução passa por continuar com novos ciclos de treinamento ou aumentar a quantidade de neurônios no

modelo, oferecendo à rede um maior grau de liberdade. Para a situação de *overfitting* o erro médio durante a fase de treinamento pode vir a ser extremamente baixo mas, ao se aplicar o conjunto de validação o erro tende a ser muito alto. A ocorrência deste fenômeno pode ser em função da existência de um grau de liberdade muito alto na rede ou o excesso de treinamento. Entre estes dois extremos – *over* e *underfitting* –, certamente existirá um modelo ótimo (LAWRENCE; GILES; TSOI, 1996, p. 30).

5.8.1. Quantidade de Camadas

Qualquer modelo de previsão assume que existe um relacionamento entre os dados de entrada, representados pelos valores passados, e os dados de saídas, referentes a valores futuros (ZHANG; PATUWO; HU, 1998, p. 36). Baseado nesta premissa, a rede neural artificial, funcionando como aproximador universal de funções, busca mapear o comportamento dos dados passados, a fim de inferir o valor futuro.

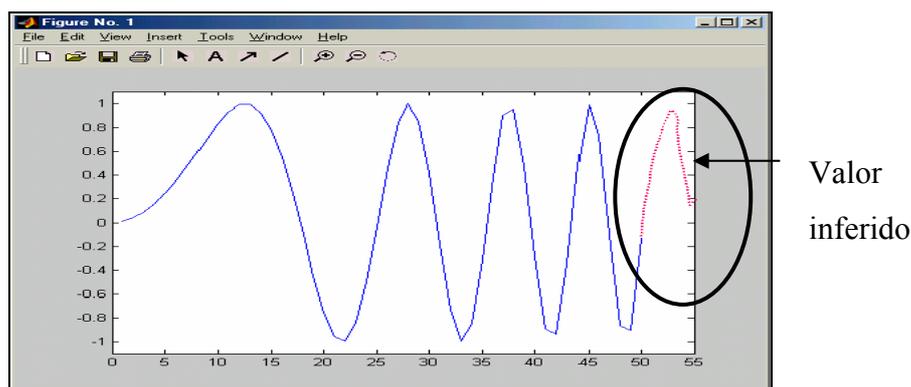


Gráfico 24: Função mapeada pela rede neural. Dentro da elipse têm-se os valores previstos.

Uma rede neural artificial do tipo *feedforward* é composta basicamente de camadas de neurônios altamente interconectados. Nesta estrutura está presente todo o conhecimento da rede, ou seja, é na matriz de pesos, modificada durante o treinamento, que a rede retém o conhecimento sobre a função de mapeamento dos dados de entrada nas saídas desejadas.

O processo realizado por cada nó da rede é definido pela combinação dos processamentos realizados pelos nós da camada anterior que estão conectados a ele. As funções implementadas definem como é realizada a divisão do espaço de decisão.

Em 1957 Kolmogorov demonstrou que funções contínuas de 3 variáveis podem ser representadas como superposições e composições de funções de 1 variável

(NASCIMENTO; YONEYAMA, 2002, p. 120).(KOVÁCS, 1996, p. 107-108). Baseado nesta teoria Cybenko, em 1989, demonstrou que redes neurais com uma camada de neurônios com funções de transferência não lineares e uma camada com funções de transferência lineares são capazes de aproximar qualquer função contínua. Acrescentando-se uma camada adicional pode-se aproximar qualquer função matemática com a acurácia desejada (BRAGA; CARVALHO; LUDEMIR, 1998, p. 54) (NASCIMENTO; YONEYAMA, 2002, p. 120-121) (ZHANG; PATUWO; HU, 1998, p. 44). Assim, ainda segundo Cybenko (ZHANG; PATUWO; HU *apud* Cybenko, 1998, p. 44):

“...that a network never needs more than two hidden layers to solve most problems including forecasting.”

5.8.2. Quantidade de Neurônios

A definição da quantidade ideal de neurônios em cada camada de uma rede neural é um assunto controverso em que os pesquisadores ainda não conseguiram chegar a um denominador comum. Várias propostas são feitas, através de artigos especializados, e, entretanto, não há consenso entre as diversas sugestões apresentadas. Regras para se limitar o número de neurônios nas camadas, visando reduzir o efeito do *overffiting*, também têm sido propostas e caem no mesmo problema.

Para citar alguns exemplos, têm-se, nos casos de apenas 1 camada escondida, as sugestões de Lippmann e Hecht-Nielsen de se usar “ $2n + 1$ ” neurônios, onde n corresponde ao número de variáveis de entrada, já Wong propõe “ $2n$ ” neurônios, Tang avaliza o uso de “ n ” como o ideal e Kang “ $n/2$ ” neurônios. Todos relacionados no artigo de Zhang (ZHANG; PATUWO; HU, 1998, p. 44).

Zhang et al (ZHANG; PATUWO; HU, 1998, p. 44) afirmam que o caminho mais comum para se determinar o número de neurônios é via experimentos ou por tentativa e erro. Cichocki segue a mesma linha e sugere a criação de redes superdimensionadas (CICHOCKI; UNBEHAUEN, 1995, p.136-137) com a subsequente eliminação dos neurônios em excesso, através do processo de poda.

5.8.3. Conexões

Redes neurais artificiais também são conhecidas como conexionistas. Tal denominação advém do fato de o modelo neural ser caracterizado por conexões entre suas camadas, a exemplo do cérebro humano.

Redes do tipo *feedforward* são, por natureza, altamente conectadas, podendo existir a possibilidade de usá-las com maior ou menor grau de conectividade. As redes com alto grau de conexões conseguem, de uma maneira geral, convergir. Entretanto, o tempo de treinamento aumenta em razão do maior número de cálculos a serem realizados. Por outro lado, uma baixa ordem de conexões (*sparse-connectivity*), onde nem sempre o neurônio da camada anterior mantém ligação com todos os neurônios da camada seguinte, pode vir a não convergir.

As conexões entre os nós de uma rede é fundamental na determinação do seu comportamento. Nas aplicações de previsão, geralmente os neurônios da camada j são totalmente conectados aos neurônios da camada $j + 1$. Estudo promovido por Duliba consistiu em adicionar conexões diretamente da camada de entrada à de saída, obtendo, segundo o autor, melhor acurácia nos resultados da previsão (ZHANG; PATUWO; HU *apud* Duliba, 1998, p. 46). Já Tang e Fishwick também investigaram os efeitos das conexões diretas entre os neurônios de uma camada k e a camada $k + 2$ não tendo chegado a nenhuma conclusão relativa a melhora de desempenho (ZHANG; PATUWO; HU *apud* Tang e Fishwick, 1998, p. 46).

5.8.4. Funções de Ativação e de Propagação

O neurônio, como elemento de processamento (EP), é responsável por acumular o produto entre as entradas e os pesos comparando o resultado com um limiar (*threshold*) e, alcançando-o, propagar este valor adiante. Em ambos os casos, ou seja, propagando o valor calculado ou não a resposta produzida pelo neurônio é significativa pois afetará a saída da rede ou os neurônios da próxima camada. Este processo é conhecido como função de transferência.

Aqui cabe comentar que não há unanimidade nas terminologias adotadas. Alguns autores separam o processo realizado no neurônio entre função de ativação e de transferência (TAFNER; XEREZ; RODRIGUES, 1995, p. 58-59), cabendo à primeira a realização do somatório do produto das entradas pelos pesos e a comparação com o

threshold e, à segunda, a aplicação da função de transferência ou propagação sobre o valor calculado no passo anterior. Outros autores consideram todo o procedimento como sendo realizado por uma única função, denominada de ativação e o somatório do produto dos pesos como sendo uma simples combinação linear (ZHANG; PATUWO; HU, 1998, p. 46) (HAYKIN, 2001, p. 36-37), (BISHOP, 1995, p. 118). Por ser de uso mais freqüente, o primeiro dos conceitos será adotado nesta dissertação.

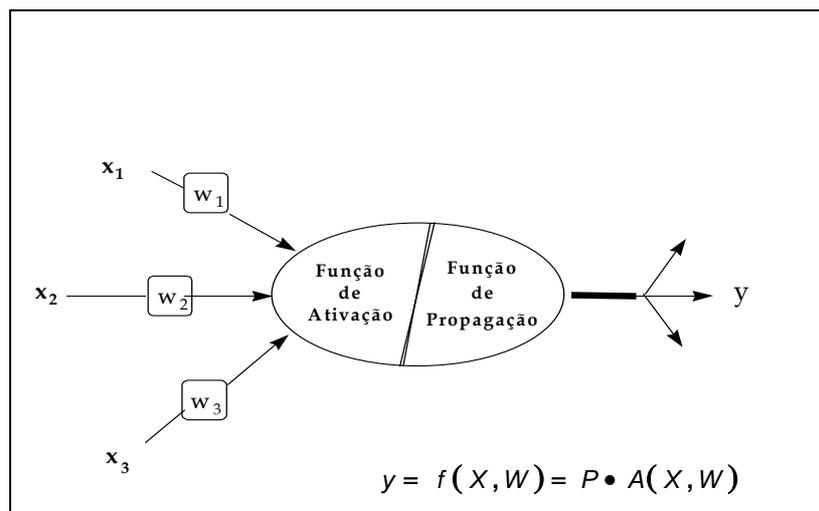


Figura 10: Representação de um neurônio artificial.

Na figura acima a função de ativação tem por atribuição ponderar a soma das entradas com os respectivos pesos (sinapses) e repassar o sinal para a saída do neurônio. Além disso, é nesta etapa em que se determina o novo estado de ativação (.) a partir das novas entradas e do estado de ativação atual. Algumas funções para se determinar o novo estado de ativação são apresentadas a seguir, sendo a função linear a mais aceita e usada atualmente.

- Linear

$$a_i(\underline{x}, \underline{w}) = \sum_{j=1}^n w_{ij} x_j \quad \text{Eq. (39)}$$

- Mahalanobis

$$a_i(\underline{x}, \underline{w}) = (\underline{x} - \underline{w}_i)^T \Omega^{-1} (\underline{x} - \underline{w}_i) \quad \text{Eq. (40)}$$

sendo Ω a matriz de covariância dos valores de saída.

- Polinomial

$$a_i(\underline{x}, \underline{w}) = \sum_{j=1}^n x_j w_{ij} \quad \text{Eq. (41)}$$

A função de propagação $\varphi(\cdot)$ determina o relacionamento entre as entradas e saídas da rede. Em geral ela é responsável pela introdução da não-linearidade nas aplicações neurais e, a princípio, qualquer função pode ser utilizada. Entretanto, na prática, somente um pequeno número delas são empregadas pois em razão dos cálculos aplicados através dos algoritmos de treinamento, há necessidade dessas funções serem limitadas, monotonicamente crescentes e deriváveis em todos os seus pontos (ZHANG; PATUWO; HU, 1998, p. 47).

Como funções de propagação, as mais encontradas na literatura são:

- Função degrau: limita a saída do neurônio a apenas dois valores (0 ou 1). Normalmente é utilizada para criar neurônios que tomem decisões binárias, como nos classificadores. É limitada (“*bounded*”), porém não é derivável;
- Função linear: não é limitada. Neurônios com esta função de propagação podem ser utilizados como aproximadores lineares;
- Função sigmoideal (logística): permite que a entrada assuma qualquer valor no intervalo $(-\infty \text{ e } \infty)$ e os comprime para o intervalo $[0, +1]$. É a função geralmente adotada em redes neurais, em virtude de ser contínua, monotônica, não linear e diferenciável em n ordens;
- Função tangente hiperbólica: mapeia a entrada dos neurônios no intervalo $[-1, +1]$. Similar a função logística sigmoideal, possibilitando, entretanto, que as saídas sejam simétricas.

É importante lembrar que a construção de uma rede neural não obriga, necessariamente, a se usar a mesma função em todas as camadas. Ao contrário, geralmente os melhores resultados são obtidos utilizando-se diferentes funções. Contudo, não se deve misturar na mesma camada funções distintas pois haverá um aumento na complexidade dos cálculos, gerando, conseqüentemente, um aumento no tempo de processamento.

Nos problemas envolvendo classificação, a função logística é a mais aplicada. Entretanto, na área de previsão utiliza-se a função linear na camada de saída. Referências sobre o trabalho de vários pesquisadores, assim como as funções por eles adotadas podem ser encontradas no artigo do Zhang et al (ZHANG; PATUWO; HU, 1998, p. 47)

5.9. DINÂMICA OU TREINAMENTO DA REDE

A segunda grande área a ser avaliada para se buscar um bom desempenho de redes neurais artificiais está ligada à dinâmica da rede (sistema cujo estado varia com o tempo) e, neste contexto, têm-se a convergência e a estabilidade global.

Dentro do modelo conexionista, a convergência está relacionada com a minimização do erro entre as saídas computadas e as desejadas e, em geral, está associada ao aprendizado supervisionado. Quanto à estabilidade, o conceito refere-se a manutenção ou retorno dos pesos sinápticos a um estado de equilíbrio, sólido e estável (FERNANDES, 1995, p. 24).

Uma peculiaridade da superfície de erro que causa impacto sobre o desempenho do algoritmo de treinamento é, justamente, a presença de mínimos locais, pois em função de características inerentes ao algoritmo de *back-propagation* pode ocorrer deste ficar preso justamente nestes pontos. Nestas situações é certo que, em outra posição da superfície haverá um outro mínimo onde a função de custo será menor. É indesejável, portanto, que o processo de treinamento termine em um mínimo local qualquer.

Mínimos locais são pontos na superfície de erro que apresentam uma solução estável, embora não sejam a saída correta. Algumas técnicas são utilizadas tanto para acelerar o algoritmo de *back-propagation* como para reduzir a incidência de mínimos locais (BRAGA; CARVALHO; LUDEMIR, 1998, p.66).

- Utilizar taxa de aprendizado decrescente;
- Adicionar nós intermediários;
- Utilizar um termo *momentum*;
- Adicionar ruído aos dados.

Abaixo tem-se uma figura representando o conceito de mínimo local e global.

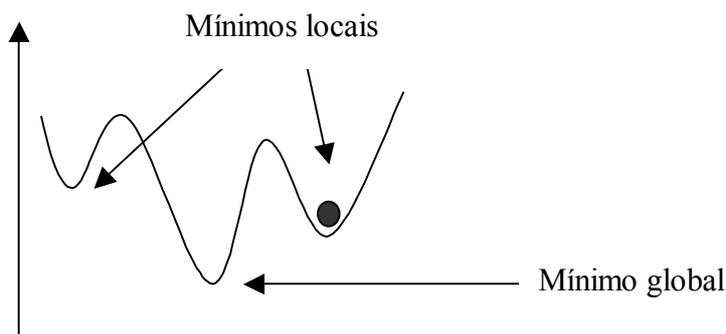


Figura 11: Representação da superfície de erro com mínimos locais e globais.

Teoricamente, uma rede do tipo multicamadas deve ser capaz de produzir uma saída desejada a partir de qualquer entrada. Na prática, entretanto, a tarefa de pré-processamento dos dados é importante para se obter uma rede eficiente pois o uso de valores sem tratamento prévio nem sempre consegue levar a rede à convergência.

5.10. REDUÇÃO DE DIMENSIONALIDADE

Dentro do escopo de pré-processamento dos dados, uma das mais importantes análises a serem feitas é a de redução de dimensionalidade. Busca-se, com isto, melhorar a performance de treinamento da rede pois, em muitas bases de dados, existem atributos dispensáveis, onde a redução de dimensionalidade não altera significativamente suas características.

5.10.1. Análise de Componentes Principais (PCA)

A análise de componentes principais consiste em tomar p variáveis X_1, X_2, \dots, X_p e achar as combinações destas que produzam índices Y_1, Y_2, \dots, Y_p , que sejam não correlacionados. A ausência de correlação é uma propriedade interessante porque mostra que os índices estão medindo dimensões diferentes do problema. Os índices são ordenados de forma que Y_1 tenha a maior variação, Y_2 a segunda maior variação, e assim por diante. Portanto, $\text{var}(Y_1) \geq \text{var}(Y_2) \geq \dots \geq \text{var}(Y_p)$, onde a $\text{var}(Y_i)$ indica a variância de Y_i no conjunto de dados considerado. Os valores de Y_i são chamados de componentes principais. Quando se realiza este tipo de análise espera-se que a variância da maioria dos índices seja tão pequena que possa ser desprezada. Nestes casos, a

variação do conjunto de dados pode ser descrita por um pequeno número de índices Y que possuam variâncias que não sejam desprezíveis. Desta forma, obtém-se alguma economia já que a variação das p variáveis X pode ser calculada através de um número menor de variáveis Y .

Deve-se ressaltar que uma análise do componente principal nem sempre resulta numa redução de um grande número de variáveis originais em um pequeno número de variáveis transformadas. Se as variáveis originais não forem correlacionadas, a análise não contribui em absolutamente nada. Os melhores resultados são obtidos quando as variáveis originais forem altamente correlacionadas, positivamente ou negativamente (EBECKEN, 2001, p. 53).

Em outras palavras, a análise de componentes principais é uma transformação linear ortogonal de um espaço p -dimensional para um espaço m -dimensional, $m < p$, onde as coordenadas dos dados no novo espaço são não correlacionadas e a maior quantidade de variância dos dados originais é preservada usando-se somente algumas poucas coordenadas no espaço transformado. O Objetivo da PCA pode ser resumido no seguinte:

- dadas p variáveis X_1, X_2, \dots, X_p , deseja-se encontrar combinações lineares que produzam vetores Y_1, Y_2, \dots, Y_p não correlacionados.

Se a maioria dos vetores apresentarem variâncias tão pequenas a ponto de serem ignoradas, a variação no conjunto de dados pode ser apropriadamente descrita pelos poucos vetores Y que retêm as maiores variâncias.

Algebricamente, os componentes principais são particularidades das combinações lineares de p variáveis aleatórias X_1, X_2, \dots, X_p . Geometricamente, estas combinações lineares representam a seleção de novo sistema de coordenadas por rotação do sistema original, com X_1, X_2, \dots, X_p como os eixos de coordenadas. Os novos eixos representam as direções com máxima variabilidade e provêm uma simples descrição da estrutura de covariância.

Os componentes principais dependem somente da matriz de covariância Σ , ou da matriz de correlação ρ dos elementos X_1, X_2, \dots, X_p . Por exemplo: seja $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ um vetor aleatório cuja matriz de covariância Σ possui os autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$.

Considere as combinações lineares adiante.

$$Y_i = e_i' X = e_{1i} X_1 + e_{2i} X_2 + \dots + e_{pi} X_p, \quad i = 1, 2, \dots, p \quad \text{Eq. (44)}$$

e $e_{11}, e_{12}, \dots, e_{ip}$ são os autovetores correspondentes.

Os componentes são extraídos na ordem do mais significativo para o menos significativo. Teoricamente o número de componentes é sempre igual ao número de variáveis. Entretanto, alguns poucos componentes são responsáveis por grande parte da informação total.

Na maioria dos problemas com que nos defrontamos o uso da primeira e segunda componente abrange 80% a 90% da população de variáveis estudadas. Dessa forma, conseguimos reduzir nosso universo de estudo sem, entretanto, perdermos muita informação.

5.10.2. Relevância das Entradas

Outra abordagem para redução da dimensionalidade é atuar no próprio espaço de entrada, analisando-se de alguma forma a relevância de cada característica

Castellano e Fanelli (CASTELLANO; FANELLI, 2000) realizaram experimentos para a seleção de dados utilizando-se de redes neurais. Segundo os autores a seleção de variáveis pode ser vista como um caso particular de poda da rede onde a eliminação dos nodos de entrada é equivalente à remoção de registros do conjunto original.

Foram feitos, nesta dissertação, alguns experimentos visando verificar a possibilidade de se reduzir a dimensão do problema com o uso de redes neurais e, para tanto, adotou-se, com algumas variações, o conceito descrito por Castellano e Fanelli.

Neste caso buscou-se verificar se todas as entradas constantes da “janela” de dados são relevantes para o modelo e, para se realizar tal análise substitui-se, uma linha de cada vez, todos os valores relativos às variáveis da matriz de treinamento por 0 (zero), apresentando a matriz modificada à rede já treinada.

Após o passo acima, comparou-se o resultado obtido com a matriz de treinamento original. Tal procedimento indica que a existência de valores muito inferiores de performance na aplicação dos dados de treinamento modificados à rede, é relevante para o modelo, não podendo excluir a linha que foi alterada. Caso contrário,

ou seja, se a quantidade de acertos de previsão, após haver sido excluída a linha, corresponder a valores próximos da performance original, indica que aquela entrada é insignificante, podendo-se retirá-la do modelo. Resumindo, a seguir tem-se os passos implementados.

Suponha uma rede com 10 elementos de entrada, 1 camada escondida e outra de saída, com um único neurônio (THOMÉ, 2001, p. 1).

1. Elimine uma das 20 entradas (substitua por 0 em toda a matriz de entradas);
2. Passe as entradas pela rede (apenas *forward*);
3. Calcule o erro médio quadrado e guarde o valor;
4. Se já tiver processado todas as entradas vá para o passo seguinte;
5. Recoloque a entrada que havia tirado e elimine a entrada seguinte, na seqüência;
6. Volte ao passo 2;
7. Normalize o vetor de erros resultante, dividindo cada termo pelo somatório de todos os termos.

Tal procedimento foi repetido para todas as 10 (dez) linhas da matriz de teste e o as experiências realizadas com a massa de dados disponível mostrou que nenhuma entrada modificada sofreu significativa redução em relação às demais, caracterizando que não iria afetar drasticamente o modelo a sua retirada. Entretanto, a performance obtida para todas as entradas e para todas as agências situou-se em intervalos regulares, caracterizando que todas as entradas têm o mesmo peso no modelo, impedindo, desta forma, a sua eliminação. Os resultados encontrados podem ser vistos no quadro abaixo.

Quadro 30: Análise da relevância das entradas para as agências 02, 06 e 13

Atrasos	Agência 02		Agência 06		Agência 13	
	Acertos	%	Acertos	%	Acertos	%
D - 1	10	6.94	4	2.78	11	7.64
D - 2	16	11.11	4	2.78	10	6.94
D - 3	16	11.11	4	2.78	9	6.25
D - 4	11	7.64	6	4.17	12	8.33
D - 5	12	8.33	6	4.17	11	7.64
D - 6	13	9.03	4	2.78	9	6.25
D - 7	17	11.81	6	4.17	11	7.64
D - 8	13	9.03	5	3.47	14	9.72
D - 9	10	6.94	4	2.78	7	4.86
D - 10	12	8.33	5	3.47	9	6.25

5.10.3. Componente Principal de Discriminação – CPD

Diferentemente do PCA onde o processo de análise considera apenas os padrões de entrada, o CPD busca selecionar variáveis no espaço de entrada com base no seu poder de discriminação no espaço de saída. Na figura 14 abaixo pode-se observar a diferença de resposta dos métodos PCA e CPD face a um mesmo problema.

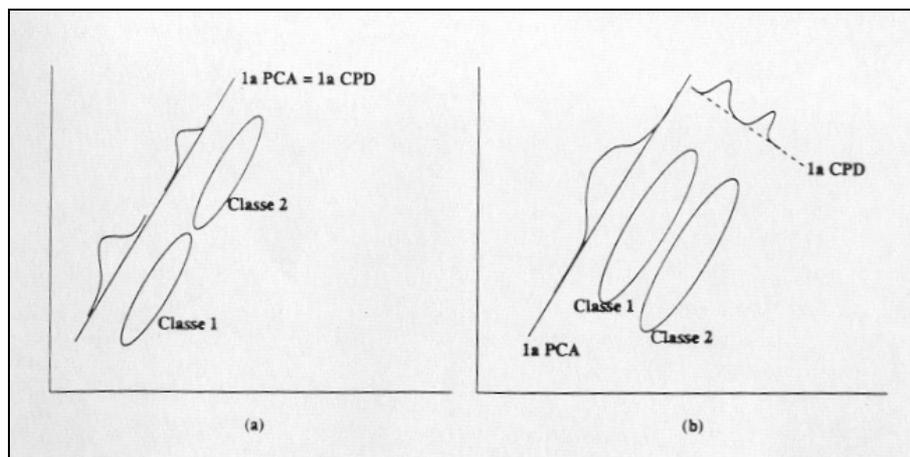


Figura 12: Representação da PCA e CPD. Na figura à esquerda tanto a PCA como a CPD conseguem discriminar as classes. Na figura (b) somente a CPD consegue definir a separação das classes.

A diferença entre os dois tipos de análise advém do fato do PCA selecionar a sua primeira componente como a direção de maior variância, não conseguindo, com isto, separar as classes. Já nos caso do CPD, o objetivo é selecionar a dimensão que ofereça a melhor separação entre as classes. Na verdade, o CPD tenta identificar certas direções privilegiadas nas quais os vetores de entrada podem ser projetados para melhor resolver o problema de classificação, com o mínimo de informação possível.

Dessa forma, ao invés de procurar por um número restrito de componentes ortogonais, que englobariam o máximo possível da variância dos dados, como é o caso do PCA, o CPD procura identificar um subconjunto de M componentes de entrada que seja suficiente para obter uma alta eficiência na classificação dos vetores N -dimensionais de entrada ($M \ll N$). Tipicamente, essa compactação leva a uma redução significativa da quantidade de processamento necessária à operação da rede neural (SOARES FILHO, 2001, p.95-96).

Para se obter este valor M treina-se a rede inicialmente com apenas um neurônio e, após a convergência, congela-se os valores dos pesos e “bias” deste neurônio, pois eles representam a componente discriminante já extraída. Após esta

etapa insere-se um outro neurônio e treina-se a rede novamente até se obter um erro médio quadrático inferior ao do treinamento anterior. Caso o erro médio quadrático se estabilize ou aumente, interrompe-se o treinamento e considera-se a quantidade de neurônios usada até aquele ponto como a ideal.

Este método de extração de componentes é conhecido como construtivo e foi adotado por Soares Filho (SOARES FILHO, 2001) em sua tese de doutorado, obtendo, segundo ele, excelentes resultados na discriminação do ruído irradiado de 4 navios. No trabalho é apresentado também o método de extração de componentes através dos resíduos do sinal de entrada. Leerink et al (LEERINK et al, 1995, p. 10) também sugere o uso desta técnica para selecionar a melhor arquitetura de uma rede neural.

5.11. REGRAS DE APRENDIZADO

Haykin descreve o processo de aprendizagem no seu livro como (HAYKIN, 2001, p. 75):

“A propriedade que é de importância primordial para uma rede neural é a sua habilidade de aprender a partir de seu ambiente e de melhorar o seu desempenho através da aprendizagem. A melhoria do desempenho ocorre com o tempo de acordo com alguma medida preestabelecida. Uma rede neural aprende acerca do seu ambiente através de um processo iterativo de ajustes aplicados a seus pesos sinápticos e níveis de polarização (bias). Idealmente, a rede se torna mais instruída sobre o seu ambiente após cada iteração do processo de aprendizagem.”

O processo de aprendizagem requer treinamento através do qual busca-se a minimização do erro existente entre o valor real e o obtido. Para tanto, várias técnicas são utilizadas. A técnica mais em uso na atualidade é a minimização do gradiente descendente implementado através do algoritmo de *back-propagation*. Por ser o algoritmo de maior aceitação no treinamento de redes neurais do tipo MLP, vários autores o descrevem em detalhes. Encontra-se no capítulo 3, seção 3.7 do livro do Cichocki (CICHOCKI; UNBEHAUEN, 1995, p.122-132) uma excelente descrição dos passos a serem adotados na implementação do algoritmo, com várias ilustrações para auxiliar o perfeito entendimento. Tem-se, também, uma pormenorizada descrição no capítulo 4 do livro do Zurada (ZURADA, 1992, p. 163-206), onde o autor apresenta

toda a evolução do algoritmo, desde a regra delta passando pela regra delta generalizada até chegar-se ao *back-propagation*. Haykin (1999, p. 161-173), Bishop (1996, p. 140-146) e Braga (BRAGA; CARVALHO; LUDEMIR, 1998, p.61-64) também desenvolvem com maior ou menor detalhe o algoritmo.

Apesar deste algoritmo ser eficiente do ponto de vista da correção dos pesos dos neurônios nas camadas escondidas, a sua convergência ocorre de forma muito lenta. Para solucionar este problema vários algoritmos de aceleração são propostos por pesquisadores, tais como o Rprop (*Resilient Back-Propagation*), o *QuickProp* etc. Está também intimamente ligado ao desempenho do algoritmo a parametrização das constantes de aprendizado e momento.

5.11.1. Taxa de Aprendizado

Embora o erro total E seja definido pela soma dos erros dos nodos de saída para todos os padrões, é assumido, sem perda de generalidade, que a minimização do erro para cada padrão individualmente levará a minimização do erro total. Assim o erro passa a ser definido pela equação abaixo (BRAGA; CARVALHO; LUDEMIR, 1998, p. 61).

$$E = \frac{1}{2} \sum_{j=1}^k (y_{d_j} - \hat{y}_j)^2 \quad \text{Eq. (45)}$$

A regra delta sugere que a variação dos pesos seja definida de acordo com o gradiente descendente do erro com relação ao peso, ou seja, que a variação do peso para um dado padrão seja definida pela equação a seguir.

$$\Delta w_{ji} \propto -\frac{\partial E}{\partial W_{ji}} \quad \text{Eq. (46)}$$

O algoritmo de aprendizagem do *back-propagation* exige que as mudanças nos pesos sejam proporcionais a $\partial E / \partial W_{ij}$ e, para se descrever uma trajetória descendente no sentido do gradiente, é necessário se usar passadas bem curtas. A constante de proporcionalidade é a taxa de aprendizado λ que, quanto maior, maiores serão as mudanças nos pesos. Existem várias heurísticas para determinação da taxa de aprendizado, porém deve-se atentar para o fato de que um valor de λ muito pequeno tornará o aprendizado muito lento e um valor grande deste poderá causar oscilações.

Abaixo tem-se a representação esquemática do neurônio artificial com as respectivas variáveis e, em seguida a fórmula de ajuste dos pesos.

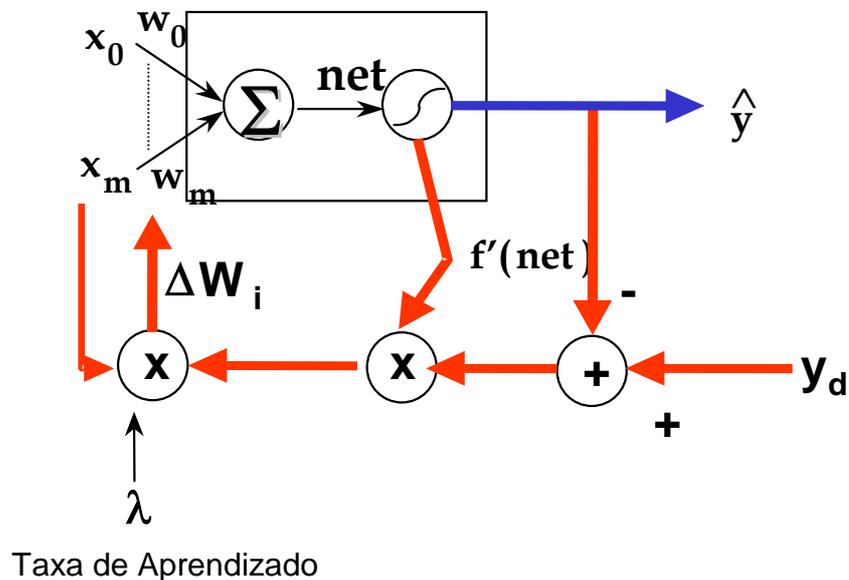


Figura 13: Neurônio artificial com parâmetros do algoritmo de back-propagation (THOMÉ, 2000, p. 16).

$$\Delta W_i = \lambda \cdot (y_d - \hat{y}) \cdot f'(net) \cdot X \quad \text{Eq. (47)}$$

onde $y_d - \hat{y}$ corresponde ao erro;

$f'(net)$ a derivada parcial da função de ativação pelo peso; e

X vetor de entrada.

Vê-se, portanto, que a variação do vetor de pesos W é diretamente proporcional à taxa de aprendizado λ e, como o que se busca é o ajuste ótimo dos pesos, depende-se, portanto, que a convergência da rede dependerá de uma boa escolha para esta constante λ . De modo geral, o valor ideal de λ irá variar de acordo com o problema a ser resolvido e não há uma constante λ única que seja a mais adequada para todos os diferentes casos. Esse é um fato bastante comum nos problemas de otimização que utilizam o método do gradiente descendente, pois as superfícies de erro podem ter um formato que dificulte o processo de convergência do algoritmo. Enquanto o método do gradiente descendente pode ser eficiente para a obtenção dos valores dos pesos que minimizem o erro, a superfície de erro frequentemente possui propriedades que fazem o treinamento convergir lentamente. (ZURADA, 1992, p. 210-211).

Existem várias heurísticas para a determinação da taxa de aprendizado, entretanto quanto menor for a escolha de λ , menores serão as variações dos pesos sinápticos da rede, de uma iteração para a outra, e mais suave será a trajetória no espaço de pesos. Esta melhoria é obtida à custa de um processo de treinamento lento. Por outro lado, um valor alto para λ provocará grandes alterações nos pesos podendo levar a rede a se tornar instável. O ideal é se ter uma taxa adaptativa, onde se torne possível alterá-la nos pontos que permitam aumentar a velocidade de treinamento. Segundo Zurada (ZURADA, 1992, p. 211), valores situados no intervalo $[10^{-3}, 10]$ foram descritos através de literatura técnica como ideais para a solução de diversos problemas. Um método simples de aumentar a taxa de aprendizagem sem no entanto instabilizar a rede é adicionando um termo de momento.

5.11.2. Momento

Rumelhart (HAYKIN *apud* Rumelhart et al, 2001, p. 196) descreveu um método para otimizar o tempo de treinamento do algoritmo de *back-propagation* sem causar uma oscilação indesejável e com melhor estabilidade do processo. Esse método que introduz o conceito de momento possibilita que a rede ignore as variações de alta frequência na superfície de erro diminuindo a probabilidade de que o processo de convergência do erro pare em um mínimo local e não global.

A utilização do momento na regra de aprendizado do *back-propagation* consiste em se fazer com que as mudanças na matriz de pesos sejam iguais a uma fração da última alteração destes pesos mais a nova alteração determinada pela regra. Assim sendo, se a alteração anterior foi realizada no sentido descendente da superfície do erro, parte da alteração atual nos pesos das conexões será realizada naquele mesmo sentido.

A fórmula para cálculo de atualização dos erros é dada conforme abaixo (THOMÉ, 2003, p. 33):

$$\Delta w_{i,j}^k = \eta \cdot x_j^k \cdot \delta_i^k \quad \text{Eq. (48)}$$

onde:

η = taxa de aprendizado;

x_j^k = j-ésima entrada da camada k;

$\delta_i^k =$ gradiente.

$$\delta_i^k = \begin{cases} P'(net_i^k).(d_i - y_i), & \text{se } k = L \\ P'(net_i^k).\sum_n w_{n,i}^{k+1}.\delta_n^{k+1}, & \text{se } k < L \end{cases} \quad \text{Eq. (49)}$$

onde:

$w_{i,j}^k =$ j-ésimo peso sináptico do i-ésimo neurônio da k-ésima camada;

$x_j^k =$ j-ésima entrada da camada k;

$net_i^k =$ valor da função de ativação do i-ésimo neurônio da camada k;

$y_i =$ i-ésima saída da rede;

$d_i =$ valor desejado para a i-ésima saída da rede;

$L =$ número de camadas.

Logo, com a adição do termo momento, a fórmula acima passa a ser escrita da seguinte forma:

$$\Delta w_{i,j}^k(t) = \eta.x_j^k.\delta_i^k + \alpha.\Delta w_{i,j}^k(t-1) \quad \text{Eq. (50)}$$

Os argumentos t e $t-1$ indicam, respectivamente, o passo atual e o anterior e α é um momento positivo, constante, definido pelo usuário. O segundo termo da equação é chamado de *termo de momento*.

A constante α situa-se no intervalo $[0, 1]$ e segundo Haykin (HAYKIN, 2001, p. 197) o valor do momento pode ser também negativo, apesar dele considerar improvável o seu uso prático. Quando o momento assume zero, a mudança nos pesos é baseada apenas na regra de aprendizado do *back-propagation* e quando assume o valor unitário a mudança atual é igual à última mudança realizada nos pesos e o gradiente é simplesmente ignorado.

Zhang et al (ZHANG; PATUWO; HU, 1988, p. 48) comentam os estudos que vários pesquisadores realizaram visando a obtenção das melhores combinações entre a taxa de aprendizado e o momento. Os resultados demonstram que os mais diversos valores podem ser aplicados. O valor ideal será definido, portanto, através dos testes realizados.

A utilização do momento é recomendada para problemas em que a convergência ocorre vagarosamente, ou para os casos em que o aprendizado não consegue chegar ao fim, por esgotamento dos recursos disponíveis.

5.12. LÓGICA NEBULOSA

Aristóteles (384 - 322 a.C.) estabeleceu um conjunto de regras para que conclusões pudessem ser aceitas logicamente. O emprego da lógica de Aristóteles levava a uma linha de raciocínio baseado em premissas e conclusões. Desde então a lógica ocidental tem sido binária, isto é, uma declaração é falsa ou verdadeira, não podendo ser ao mesmo tempo parcialmente verdadeira ou falsa (TAKEMURA, 2003).

A lógica de Aristóteles trata as sentenças classificando-as como verdadeiras ou falsas. Não obstante, muitas das experiências humanas não podem ser classificadas simplesmente como verdadeiras ou falsas, branco ou preto. Este tipo de resposta, ou seja, um sim ou não a estas questões é, na maioria das vezes, incompleto. Na verdade, entre a certeza de ser e a certeza de não ser, existem infinitos graus de incerteza. Esta imperfeição intrínseca à informação representada numa linguagem natural, tem sido tratada matematicamente no passado com o uso da teoria das probabilidades.

A Lógica Nebulosa (*Fuzzy Logic*) viola as suposições Aristotélicas. O conceito de dualidade, estabelecendo que algo pode e deve coexistir com o seu oposto, faz a lógica nebulosa, também conhecida por lógica difusa, parecer natural. Além disso, a Lógica Nebulosa, com base na teoria dos Conjuntos Nebulosos, tem se mostrado mais adequada para tratar imperfeições da informação do que a teoria das probabilidades. De forma mais objetiva e preliminar, podemos definir Lógica Nebulosa como sendo uma ferramenta capaz de capturar informações vagas, em geral descritas em uma linguagem natural e convertê-las para um formato numérico, de fácil manipulação pelos computadores de hoje em dia.

A lógica nebulosa é utilizada para representar modelos de raciocínio impreciso, que possuem um papel essencial na notável habilidade humana para tomar decisões racionais em ambientes de incertezas e imprecisões. Com o advento dessa teoria, obteve-se uma ferramenta robusta para se representar várias facetas do conhecimento humano.

Tradicionalmente, na teoria dos conjuntos, um elemento *pertence* ou *não pertence* a um conjunto. Os conjuntos nebulosos, uma generalização de um conjunto ordinário, permitem a definição de um *grau de pertinência* para cada elemento, isto é, um número real no intervalo $[0,1]$. Neste caso, se o grau é *zero*, o elemento não pertence ao conjunto e, se é 1, o elemento *pertence* totalmente ao conjunto.

Portanto, um elemento pode pertencer parcialmente a um conjunto nebuloso. Podemos citar, por exemplo, o conjunto de pessoas "*jovens*". Quando uma pessoa não é mais jovem? A definição de um conjunto nebuloso pode considerar, por exemplo, uma pessoa de 20 anos como "*90% jovem*", enquanto que alguém de 60 anos seria, apenas "*10% jovem*".

A teoria nebulosa surgiu com a publicação do artigo intitulado "*Fuzzy Sets*" por Lotfi A. Zadeh, em 1965. Neste trabalho, ele introduziu o conceito de conjunto nebuloso, cujas fronteiras não são bem demarcadas ou precisas, em contraste com os conjuntos tradicionais, nitidamente definidos e que possuem fronteiras precisas. (ZADEH, 1965, p.338-353). Este trabalho é apresentado, também, como uma abordagem alternativa para o tratamento de incertezas pois ele mostra que em muitos aspectos essa teoria se parece com a teoria da probabilidade – o uso do intervalo unitário e o fato de que ambos os conceitos descrevem incertezas.

5.12.1. Teoria dos Conjuntos Tradicionais

A teoria dos conjuntos nebulosos é, em grande parte, uma extensão da teoria dos conjuntos tradicionais. Dessa forma cabe apresentar uma revisão dos conceitos desta teoria.

Existem três formas através do qual um conjunto A pode ser definido:

- a. Um conjunto A cujo os membros são a_1, a_2, a_3 é geralmente definido por:

$$A = \{ a_1, a_2, a_3 \} \quad \text{Eq. (51)}$$

Este tipo de definição é aplicado somente a conjuntos finitos.

- a. Um conjunto também pode ser representado por:

$$A = \{x|P(x)\} \quad \text{Eq. (52)}$$

onde o símbolo $|$ denota a frase “tal que,” e $P(x)$ designa a proposição da forma “ x tem a propriedade P ”. Isto é, A é definido por esta notação como o conjunto de todos elementos de X para qual a proposição $P(x)$ é verdadeira.

- a. Por fim, um conjunto pode ser definido por uma função, geralmente chamada de *função característica*, que declara quais elementos de X são membros do conjunto e quais não são. Então, nestes casos, um conjunto A é definido por sua função característica, γ_A representada a seguir.

$$\gamma_A(x) = \begin{cases} 1 & \text{para } x \in A \\ 0 & \text{para } x \notin A \end{cases} \quad \text{Eq. (53)}$$

Caso um conjunto não contenha nenhum membro ele é chamado de conjunto vazio e representado por \emptyset . Alguns conjuntos tradicionais podem ser vistos no quadro adiante.

Quadro 31: Exemplos de conjuntos tradicionais.

Conjunto	Elementos	Descrição
Z	$\{\dots, -2, -1, 0, 1, 2, \dots\}$	Conjunto dos números inteiros
N	$\{1, 2, 3, \dots\}$	Conjunto dos números naturais
Z_0	$\{0, 1, 2, \dots\}$	Conjunto dos números inteiros não negativos
R		Conjunto dos números reais
R^+		Conjunto dos números reais não negativos

Dentre as operações utilizadas na teoria dos Conjuntos Tradicionais pode-se destacar a *união* representado por \cup , *intersecção* representado por \cap e o *complemento* representado por \neg .

A união dos conjuntos A e B é denotada por:

$$A \cup B = \{x | x \in A \text{ ou } x \in B\} \quad \text{Eq. (54)}$$

A intersecção de dois conjuntos A e B é denotada por:

$$A \cap B = \{x | x \in A \text{ e } x \in B\} \quad \text{Eq. (55)}$$

O complemento representa os elementos de um universo que não fazem parte de um conjunto:

$$\neg A = \{x | x \in X \text{ e } x \notin A\} \quad \text{Eq. (56)}$$

As operações de união, intersecção e complemento de conjuntos tradicionais possuem várias propriedades, algumas delas compiladas no quadro abaixo, onde A , B e C são conjuntos tradicionais definidos sobre um universo X .

Quadro 32: Propriedades Fundamentais das operações sobre Conjuntos Tradicionais

Nº	Propriedade	Representação
01	Absorção	$A \cup (A \cap B) = A$ $A \cap (A \cup B) = A$
02	Absorção por X e \emptyset	$A \cup X = X$ $A \cap \emptyset = \emptyset$
03	Associatividade	$A \cup (B \cup C) = (A \cup B) \cup C$ $A \cap (B \cap C) = (A \cap B) \cap C$
04	Comutatividade	$A \cup B = B \cup A$ $A \cap B = B \cap A$
05	Distributividade	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
06	Idempotência	$A \cup A = A$ $A \cap A = A$
07	Identidade	$A \cup \emptyset = A$ $A \cap X = A$
08	Involução	$\neg \neg A = A$
09	Lei de Contradição	$A \cap \neg A = \emptyset$
10	Lei De Morgan	$\neg (A \cup B) = \neg A \cap \neg B$ $\neg (A \cap B) = \neg A \cup \neg B$
11	Lei do Meio Excluído	$A \cup \neg A = X$

5.12.2. Teoria dos Conjuntos Nebulosos

Um conjunto nebuloso é definido em um universo de discurso (conjunto base) X , e caracterizado pela sua função de pertinência:

$$A : X \rightarrow [0,1] \quad \text{Eq. (57)}$$

onde $A(x)$ representa o grau com que x pertence a A e expressa a extensão com que x se enquadra na categoria representada por A .

Uma função de pertinência particular pode ser visualizada por meio da equação abaixo. Verifica-se que esta função é triangular e as variáveis a , b e c são parâmetros da função.

$$\mu(x) = \begin{cases} \frac{x-a}{b-a} & \text{se } x \in [a, b) \\ \frac{c-x}{c-b} & \text{se } x \in [b, c] \\ 0 & \text{caso contrário} \end{cases} \quad \text{Eq. (58)}$$

Conforme definido anteriormente, a teoria dos conjuntos nebulosos é uma extensão da teoria dos conjuntos tradicionais. Assim, as principais operações e relações entre conjuntos nebulosos são definidas como extensão das operações e relações tradicionais, como pode ser visto no quadro abaixo, onde A e B denotam conjuntos nebulosos sobre um conjunto base X e $A(x)$ e $B(x)$ representam os graus de pertinência de x nos conjuntos nebulosos A e B respectivamente.

Quadro 33: Operações e relações com conjuntos nebulosos.

Nº	Operação	Representação	Natureza
1	Complemento	$\neg A(x) = 1 - A(x)$	Operação
2	Diferença	$(A \neq B)$ se $A(x) \neq B(x)$ para pelo menos um elemento de $x \in X$	Relação
3	Igualdade	$(A = B)$ se $A(x) = B(x)$ para todo $x \in X$	Relação
4	Inclusão	$(A \subseteq B)$ se $A(x) \leq B(x)$ para todo $x \in X$	Relação
5	Intersecção	$A \cap B = A(x) \cap B(x) = \min [A(x), B(x)]$	Operação
6	União	$A \cup B = A(x) \cup B(x) = \max [A(x), B(x)]$	Operação

Além das operações e das relações os conjuntos nebulosos possuem algumas características especiais. Entre tais características encontram-se: corte α (*cut*), conjunto de níveis, suporte, altura e normalização. A seguir tais características descritas acima serão apresentadas de forma sintética, supondo que A é um conjunto nebuloso sobre o conjunto base X .

Corte α

O corte α (${}^\alpha A$) de um conjunto nebuloso A corresponde ao conjunto tradicional que contém todos os elementos do conjunto universo X com grau de pertinência em A maior ou igual a α , enquanto que o corte α forte (${}^{\alpha+} A$) contém todos os elementos em um conjunto universo X com grau maior que α , onde $\alpha \in [0, 1]$.

$${}^{\alpha}A = \{x \in X \mid A(x) \geq \alpha\} \quad \text{Eq. (59)}$$

$${}^{\alpha+}A = \{x \in X \mid A(x) > \alpha\} \quad \text{Eq. (60)}$$

Conjunto de Níveis

O conjunto de níveis (Λ) de um conjunto nebuloso A corresponde a um conjunto que contém todos os valores $\alpha \in [0,1]$ e que representam cortes α de A distintos. O conjunto de níveis do conjunto nebuloso A é representado formalmente por:

$$\Lambda A = \{\alpha \mid A(x) = \alpha \text{ para algum } x \in X\} \quad \text{Eq. (61)}$$

Suporte

O suporte de um conjunto nebuloso A , em um conjunto universo X , é o conjunto tradicional que contém todos os elementos de X que possuem grau de pertinência diferente de zero em A . Claramente, o suporte de A é exatamente o mesmo que o corte α forte de A para $\alpha = 0$. Vários símbolos especiais costumam ser usados para representar o suporte de um conjunto, tais como: $S(A)$ ou $sup(A)$.

$$sup(A) = \{x \in X \mid A(x) > 0\} \quad \text{Eq. (62)}$$

Altura

A Altura (h) de um conjunto nebuloso A corresponde ao seu maior grau de pertinência, entre todos os elementos do conjunto.

$$h(A) = \max_{x \in X} A(x) \quad \text{Eq. (63)}$$

Normalização

Um conjunto nebuloso A é chamado de *Normal* quando a sua altura é igual a 1, ou seja, pelo menos um grau de pertinência, dos elementos do conjunto, possui valor máximo, enquanto que os conjuntos que não possuem altura igual a um são chamados de *subnormal*. Portanto:

A é dito normal se $h(A) = 1$

A é dito subnormal se $h(A) < 1$

Caso um conjunto nebuloso possua apenas um elemento com grau de pertinência igual a um, este elemento é denominado protótipo do conjunto. Um conjunto nebuloso não normalizado pode ser normalizado por meio da divisão dos graus de pertinência de cada elemento, pelo maior grau de pertinência encontrado no conjunto.

5.12.3. Sistemas Nebulosos

Um sistema nebuloso representa uma função (geralmente não-linear) entre duas ou mais variáveis através da descrição lingüística da relação entre elas. O conhecimento para esta descrição pode envolver conceitos vagos e/ou imprecisos, representados por sentenças nebulosas do tipo:

$$x \text{ é } A \quad \text{Eq. (64)}$$

Um sistema nebuloso é então um conjunto de regras do tipo se-então que mapeia entradas em saídas. Estas regras definem regiões no espaço entrada x saída ($X \times Y$). Um sistema nebuloso $F: X \rightarrow Y$ aproxima a função $f: X \rightarrow Y$ cobrindo seu gráfico com fragmentos ou remendos que se sobrepõem (*patches*). Cada regra determina um remendo. A aproximação da função melhora a medida que os fragmentos aumentam em número e diminuem em tamanho como é mostrado na figura abaixo.

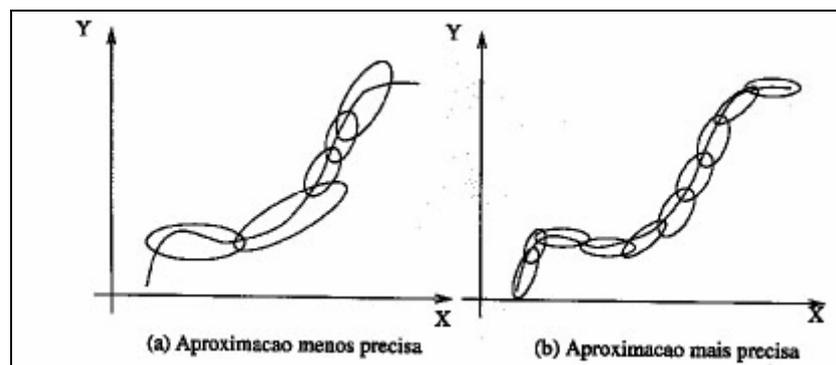


Figura 14: Gráficos mostrando como aumentar a precisão da aproximação nebulosa.

5.12.4. Regras Nebulosas

Uma regra nebulosa é uma sentença nebulosa que expressa o conhecimento da relação entre duas ou mais variáveis através de uma relação entre os conjuntos

nebulosos associados a cada sentença. Por exemplo, o conjunto nebuloso A pode ser ligado ao conjunto nebuloso B pela regra $A \rightarrow B$, escrita como:

$$\text{Se } x \text{ é } A \text{ então } y \text{ é } B \quad \text{Eq. (65)}$$

Um modelo lingüístico relacionando as variáveis x e y pode ser escrito como um conjunto de N regras que relacionam os termos $A_i, i = 1 \dots n$ e $B_j, j = 1 \dots m$, por regras $A_i \rightarrow B_j$. Cada combinação (i,j) representa uma regra k escrita como:

$$\text{Se } x \text{ é } A_i \text{ então } y \text{ é } B_j, k = 1 \dots N \quad \text{Eq. (66)}$$

5.13. CONCLUSÕES DO CAPÍTULO

Neste capítulo apresentou-se algumas das técnicas clássicas (amortecimento exponencial e ARIMA) e não clássicas (redes neurais artificiais e lógica nebulosa) usadas para previsão de séries temporais.

Geralmente o uso de redes neurais artificiais vem apresentando melhores resultados que qualquer outro modelo, pois o fato delas mapearem sistemas não lineares as tornam úteis para séries complexas. Por este motivo, a seção sobre RNA foi a mais explorada.

A literatura aponta as redes recorrentes como sendo as melhores para tratar de problemas cuja variável tempo esteja presente e este foi o motivo da inclusão do modelo *Elman*, que faz parte deste grupo de redes.

Entretanto, fazendo-se algumas adaptações, pode-se simular uma rede do tipo MLP (*MultiLayer Perceptron*) para trabalhar com atraso no tempo. Para tanto, uma das técnicas utilizadas é criando-se janelas de tempo e, para tanto, basta preparar os dados de entrada de forma adequada. Assim sendo, por ser uma rede de extrema importância e utilização, optou-se pela realização de experimentos utilizando-a.

Foi mostrado, também, uma análise sobre as técnicas de otimização de RNAs utilizadas na dissertação. Trata-se de matéria importante para o desenvolvimento de modelos neurais mas pouco abordada na literatura. Geralmente encontra-se alguma referência, por vezes bem superficial, sobre os conceitos e sua utilização mas poucos autores destacam esta área que apesar de trabalhosa e nem sempre gratificante, é, muitas vezes, responsável pelo sucesso do modelo ou, em última análise, pelo ganho computacional envolvido na fase de treinamento da rede. De qualquer forma, a maioria

dos trabalhos sempre informam algum tipo de tratamento que foi aplicado aos dados a fim de melhorar a performance da rede ou facilitar sua convergência.

Apresentou-se, por último, os conceitos de lógica nebulosa pois este método também foi utilizado na modelagem da série.

6. ABORDAGEM CLÁSSICA DAS SÉRIES EM ESTUDO

Neste capítulo discorre-se inicialmente sobre as componentes de uma série temporal para, na 2^a. seção, apresentar graficamente as diversas componentes envolvidas e analisá-las. No item 3 têm-se os modelos aplicados e os resultados de previsão. Na seção 4 têm-se as conclusões.

6.1. COMPONENTES DE UMA SÉRIE TEMPORAL SEGUNDO A ABORDAGEM CLÁSSICA

No método clássico de decomposição de séries temporais procura-se isolar as componentes tendência, variação cíclica, variações sazonais e variações irregulares dos dados, a fim de se analisar cada componente separadamente. A análise de tais dados tem por objetivo determinar se eles apresentam algum padrão não aleatório, pois, na maioria das vezes o que se deseja é localizar esses padrões não aleatórios e usá-los para previsões referentes ao futuro (STEVENSON, 1986, p. 412-413). Após isolar estas componentes, o estudo consiste em descrever, matematicamente, as características existentes na série (SPIEGEL, 1961, p. 471).

Precedendo a modelagem matemática, a avaliação de uma série temporal passa, primeiramente, pela construção de gráficos, pois, através deles, pode-se verificar a sua evolução ao longo do tempo, se existe algum tipo de tendência, se existe ou não sazonalidade, se ocorrem *outliers* etc.

6.2. DECOMPOSIÇÃO GRÁFICA DAS COMPONENTES

Para avaliar as componentes das séries utilizadas nesta dissertação, foram criados gráficos representativos do seu movimento. Para cada agência foi plotada a série completa e, logo após, um gráfico decomposto em tendência/ciclo e irregularidade. Primeiramente tratou-se a série com *outliers* e, em seguida, sem *outliers*, ou seja, promoveu-se as mesmas análises para o conjunto de dados original e o modificado. Neste último caso – modificado –, os valores superiores a 2 desvios padrões foram trazidos para este patamar. Cabe destacar que o modelo adotado foi o aditivo e se

considerou a sazonalidade como sendo de 5 dias, pois foi o período em que se obteve a maior correlação entre os dados.

Os gráficos relativos à série da agência 02 encontram-se a seguir.

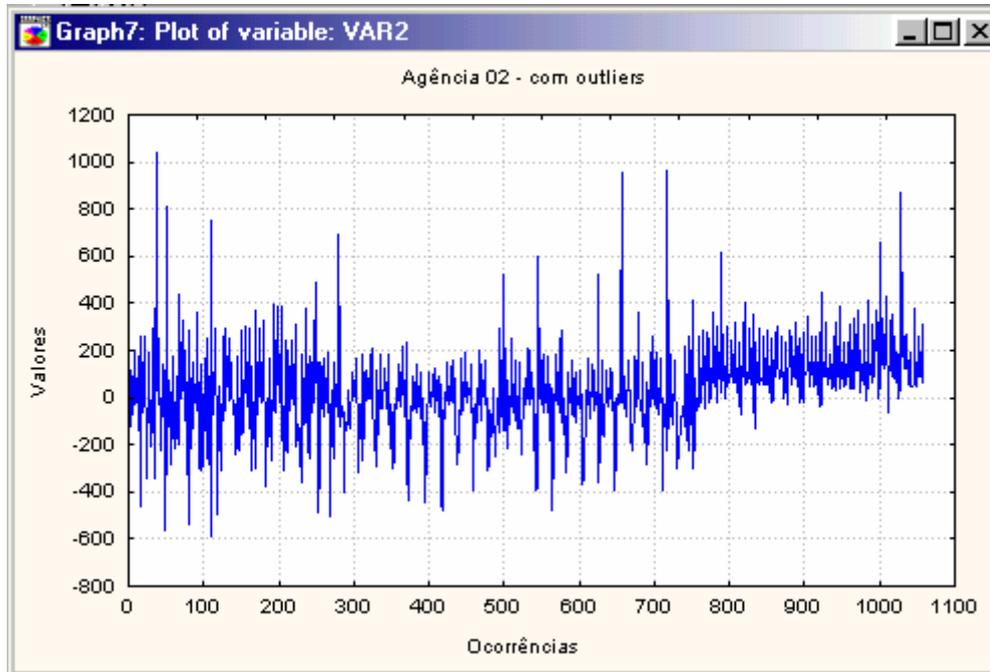


Gráfico 25: Agência 02 (com outliers). Série completa.

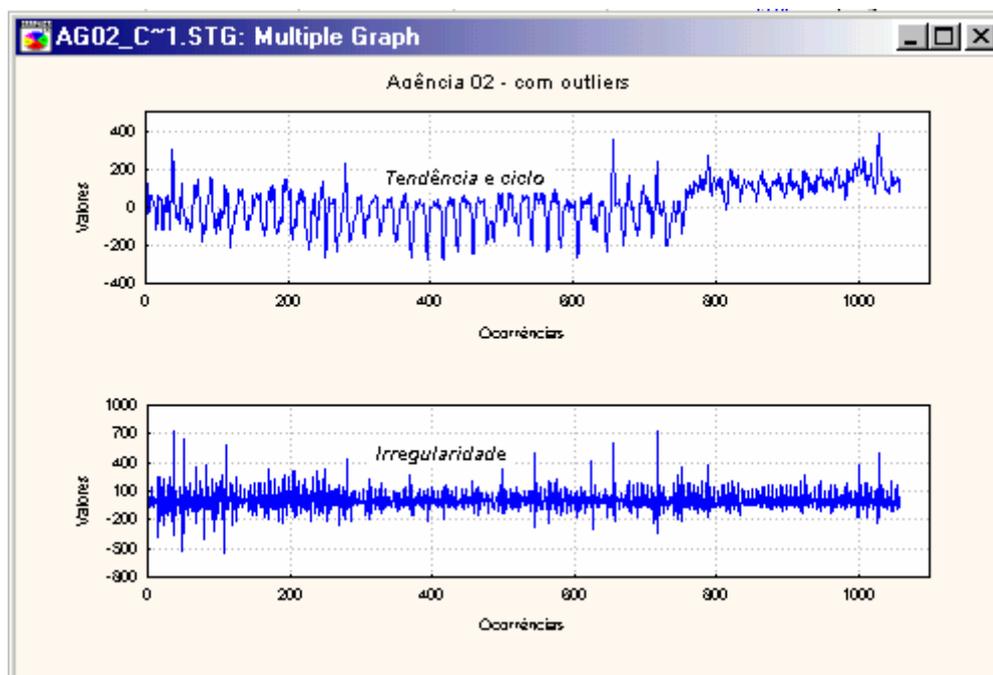


Gráfico 26: Agência 02 (com outliers) – Parte superior representando a tendência e ciclo da série e na parte inferior a irregularidade.

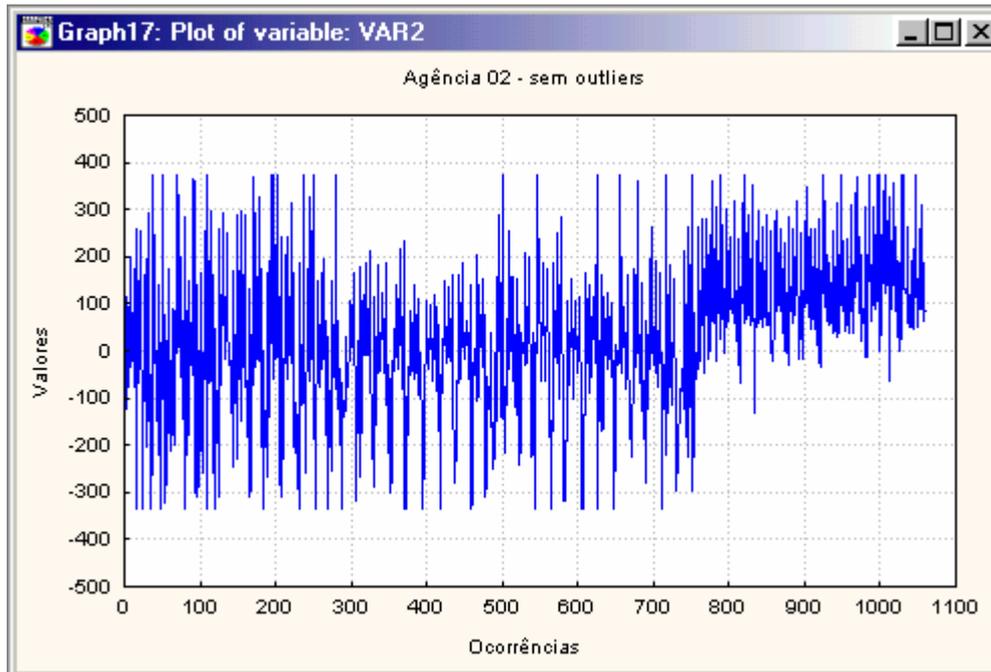


Gráfico 27: Agência 02 (sem outliers). Série completa.

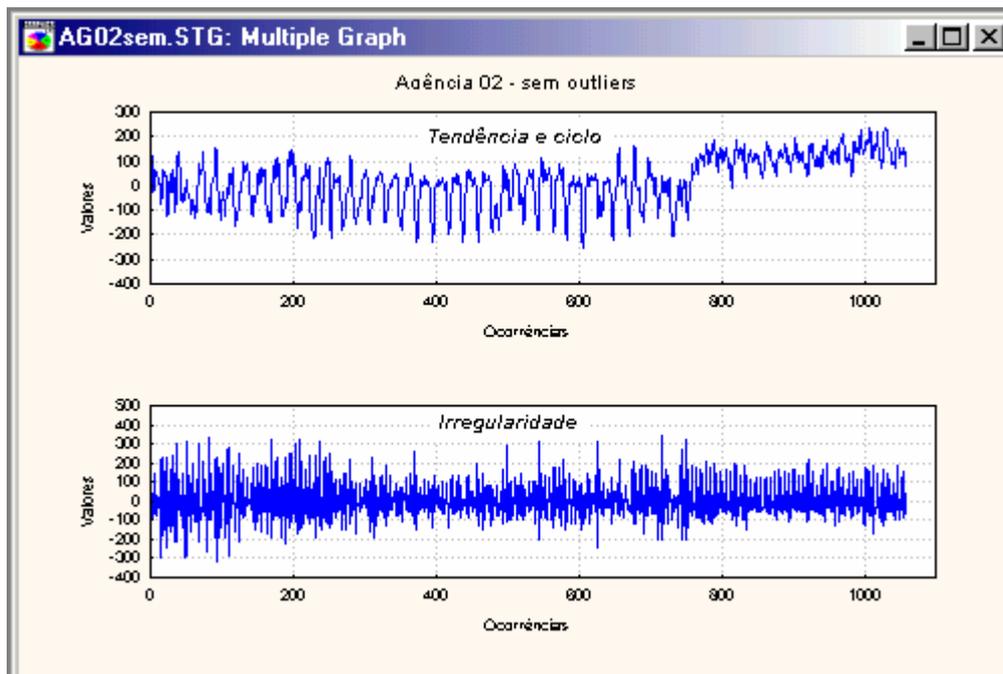


Gráfico 28: Agência 02 (sem outliers) – Parte superior representando a tendência e ciclo da série e na parte inferior a irregularidade.

Visualmente não se pode concluir muito sobre as componentes tendência e ciclo, pois não há nos gráficos nenhum desses elementos com presença muito marcante. Por outro lado, a componente irregularidade, que é formada pelo resíduo resultante da

decomposição da série após a retirada das componentes sazonalidade, tendência e ciclicidade, está presente e com valores mínimos, máximos e de desvio padrão superiores às demais componentes, conforme pode ser visto no quadro abaixo.

Quadro 34: Agência 02. Estatística das componentes da série.

	Agência 02			
	com outliers		sem outliers	
	Tendência e ciclo	Irregularidade	Tendência e ciclo	Irregularidade
Desvio padrão	106,23	119,22	98,54	99,66
Valor mínimo	-272,98	-538,55	-251,42	-318,85
Valor máximo	385,74	736,48	235,38	340,60

Pelo fato de ser, geralmente, residual, a irregularidade tende a ser de pequena amplitude e, freqüentemente, seguindo o padrão de uma distribuição normal, isto é, aquela na qual os pequenos desvios ocorrem com grande freqüência e os grandes com freqüência pequena (SPIEGEL, 1961, p. 476). Não é o que se observa. A amplitude da componente aleatória é maior que a da tendência/ciclo. Este fato induz a se acreditar que a forte presença de irregularidade na série poderá vir a dificultar a construção de um modelo de previsão eficaz.

Abaixo tem-se os histogramas das componentes irregularidade para os dados da agência 02, com e sem outliers.

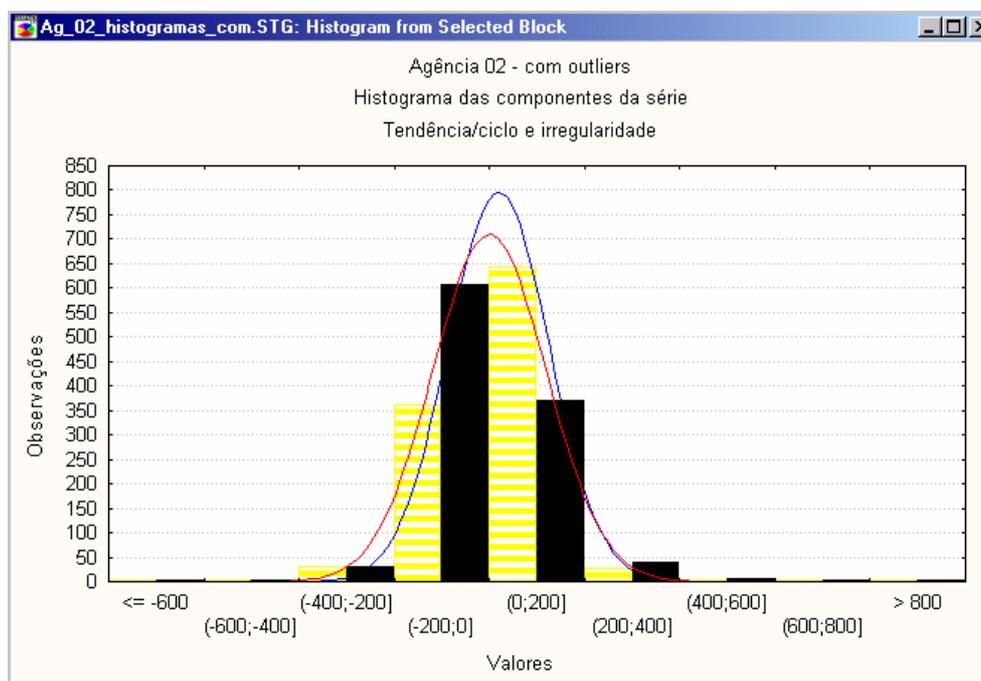


Gráfico 29: Agência 02 (com outliers) – Histograma referente às componentes tendência/ciclo (retângulo tracejado) e irregularidade (retângulo sólido).

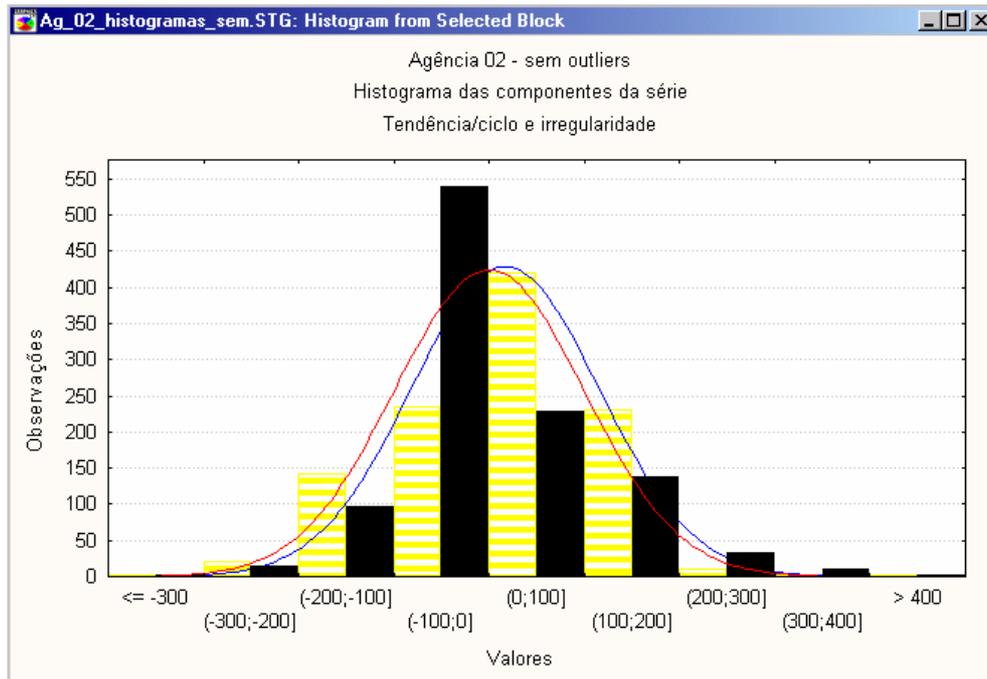


Gráfico 30: Agência 02 (sem outliers) – Histograma referente às componentes tendência/ciclo (retângulo tracejado) e irregularidade (retângulo sólido).

Em continuidade à análise gráfica das séries, a seguir têm-se os gráficos relativos à agência 06.

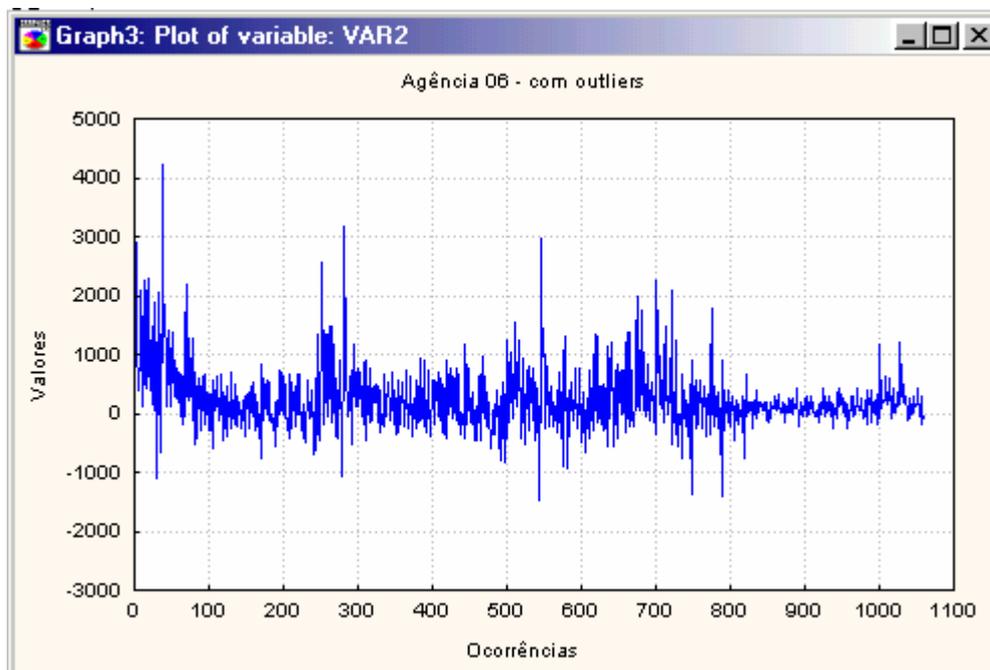


Gráfico 31: Agência 06 (com outliers). Série completa.

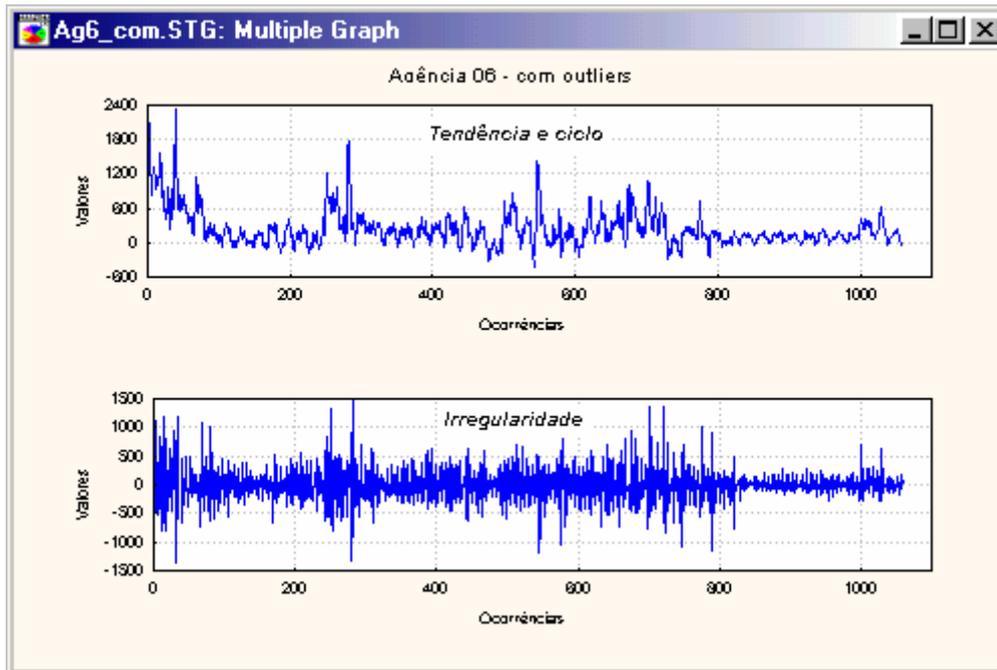


Gráfico 32: Agência 06 (com outliers) – Parte superior representando a tendência e ciclo da série e na parte inferior a irregularidade.

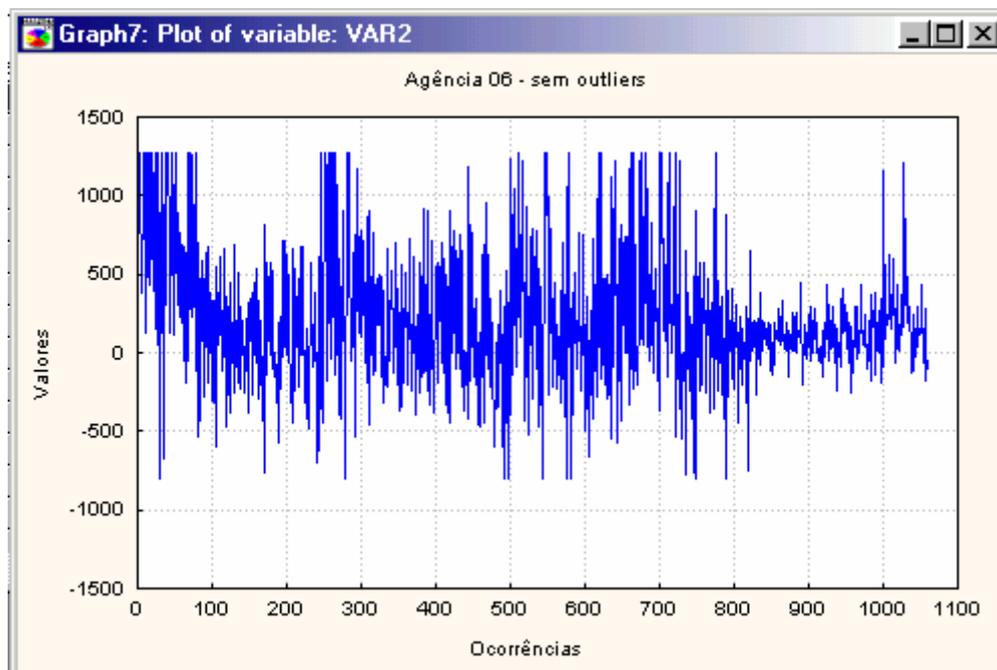


Gráfico 33: Agência 06 (sem outliers). Série completa.

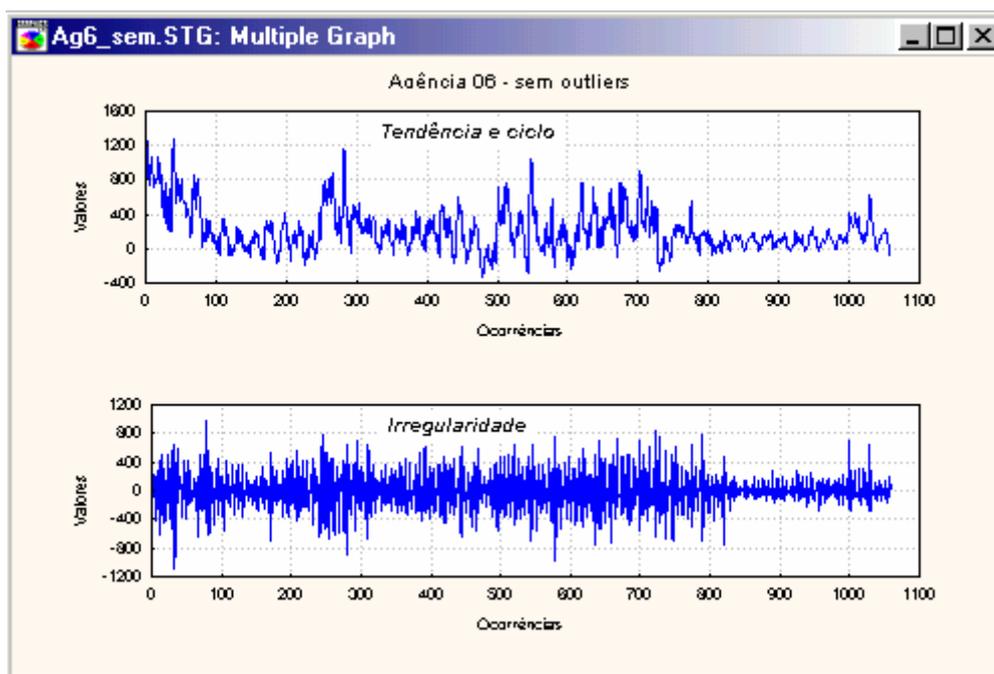


Gráfico 34: Agência 06 (sem outliers) – Parte superior representando a tendência e ciclo da série e na parte inferior a irregularidade.

Similar à agência 02, a agência 06 apresenta características de aleatoriedade bem evidentes nos seus gráficos, tendo, inclusive, grandes amplitudes. Pode-se observar, também, uma certa ciclicidade nos gráficos, relativa a períodos de 1 ano (aproximadamente 250 ocorrências).

Após a decomposição da série verifica-se que a componente irregularidade está presente com valores maiores que as demais. Esta agência também possui a mesma característica da agência anterior, ou seja, a amplitude da irregularidade é alta. Provavelmente também será uma série difícil de se modelar. Abaixo encontram-se os resultados estatísticos das componentes.

Quadro 35: Agência 06. Estatística das componentes da série

	Agência 06			
	com outliers		sem outliers	
	Tendência e ciclo	Irregularidade	Tendência e ciclo	Irregularidade
Desvio padrão	324,07	330,57	250,55	275,10
Valor mínimo	-652,50	-1.508,18	-328,20	-1.082,76
Valor máximo	2.424,10	1.854,92	1.282,90	989,07

A seguir têm-se os respectivos histogramas.

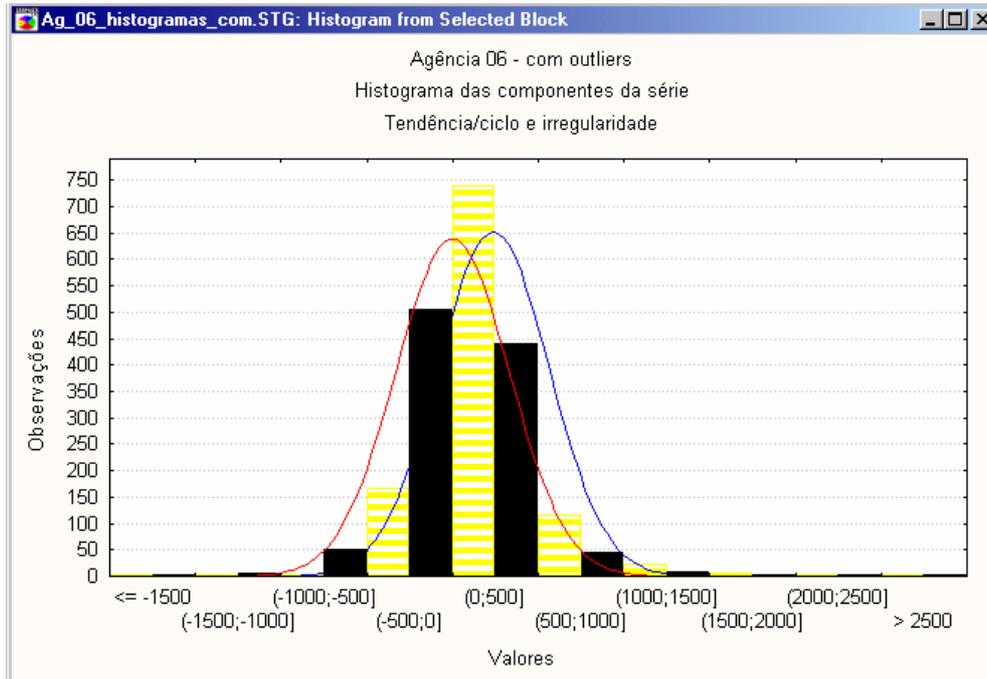


Gráfico 35: Agência 06 (com outliers) – Histograma referente às componentes tendência/ciclo (retângulo tracejado) e irregularidade (retângulo sólido).

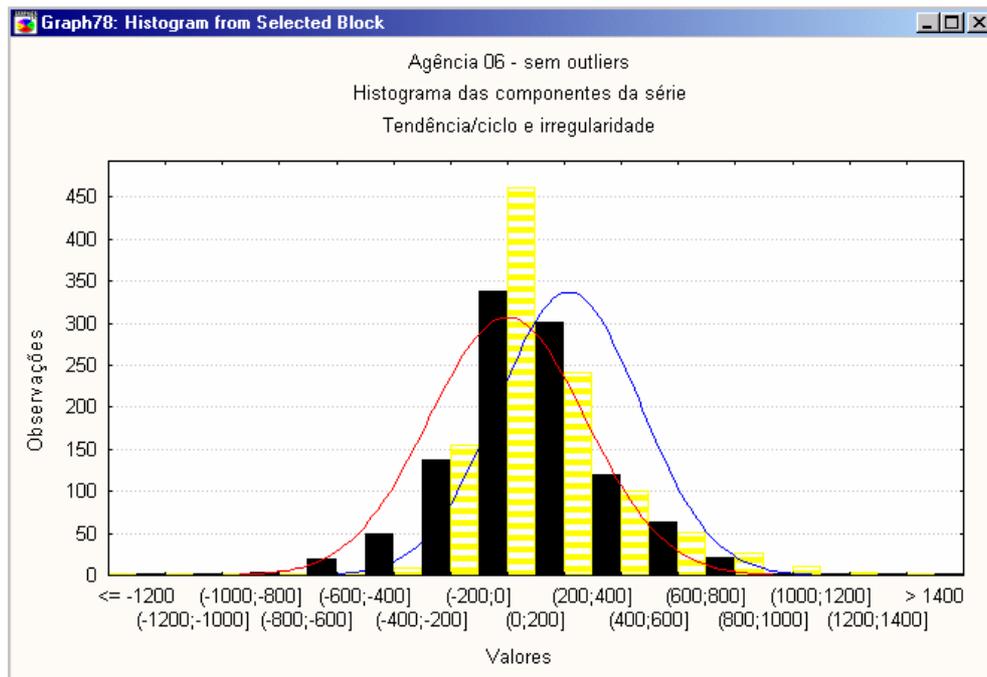


Gráfico 36: Agência 06 (sem outliers) – Histograma referente às componentes tendência/ciclo (retângulo tracejado) e irregularidade (retângulo sólido).

Finalizando, têm-se os gráficos da agência 13.

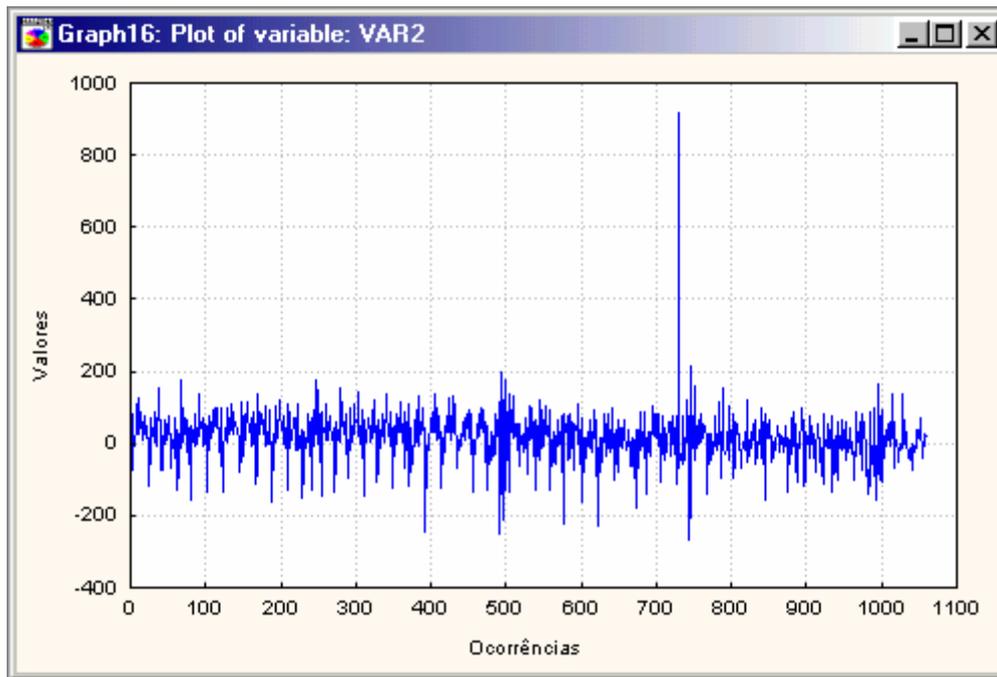


Gráfico 37: Agência 13 (com outliers). Série completa.

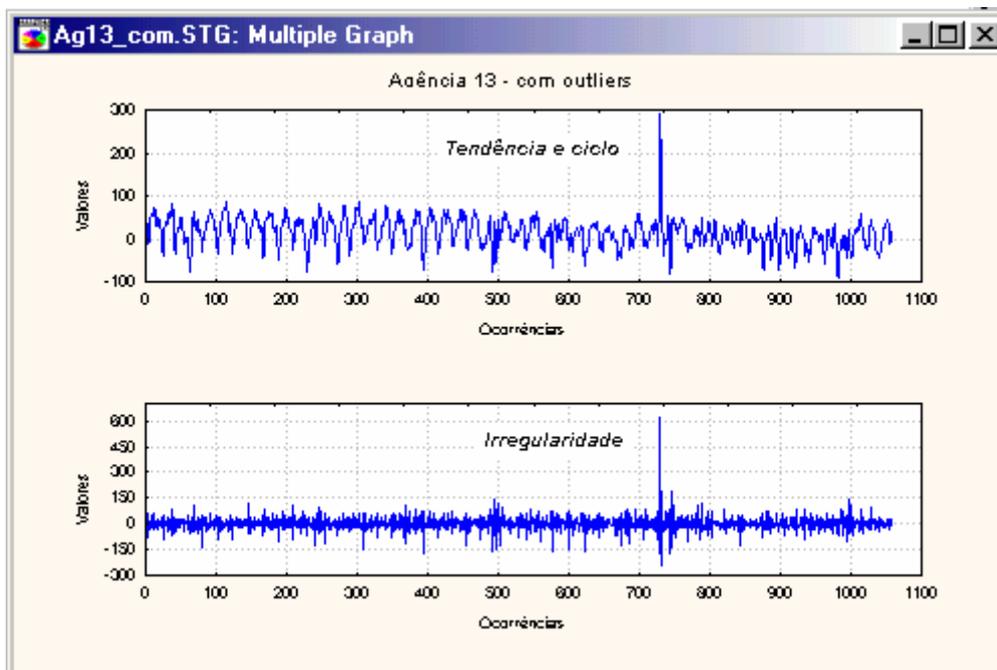


Gráfico 38: Agência 13 (com outliers) – Parte superior representando a tendência e ciclo da série e na parte inferior a irregularidade.

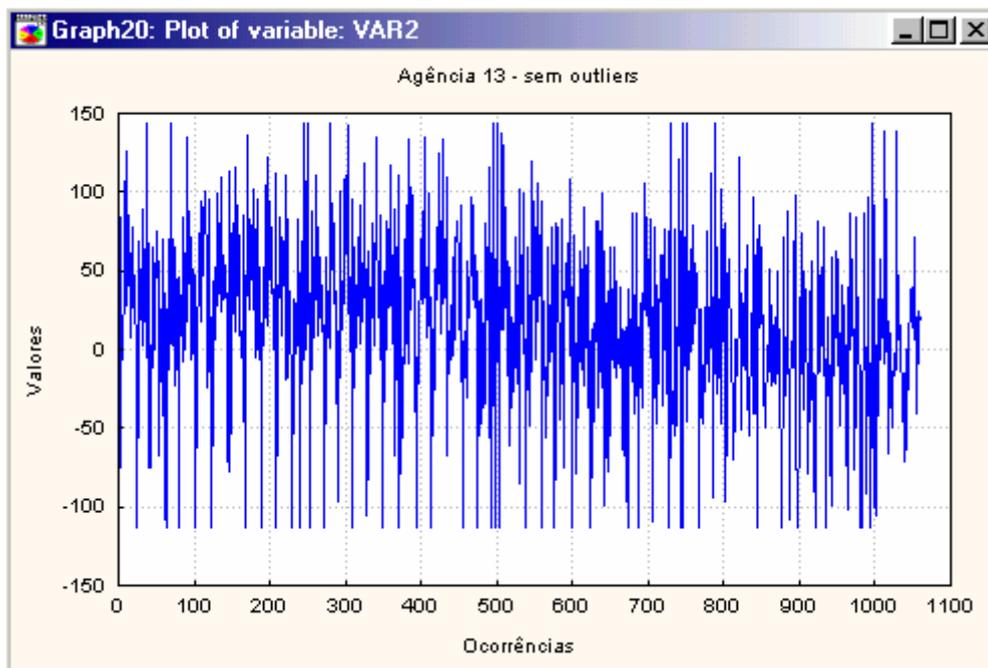


Gráfico 39: Agência 13 (sem outliers). Série completa.

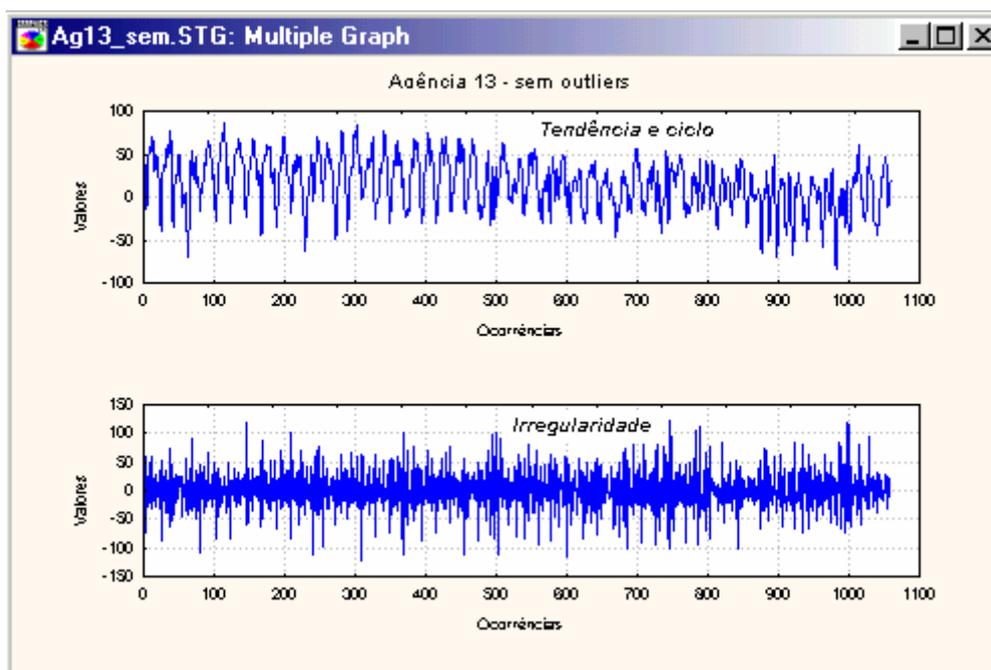


Gráfico 40: Agência 13 (sem outliers) – Parte superior representando a tendência e ciclo da série e na parte inferior a irregularidade.

A análise gráfica da agência 13 também se assemelha às das demais filiais pois a componente irregularidade é bem evidente nos gráficos. Sua amplitude, apesar de menor do que as das outras, é muito grande se comparada com o seu movimento.

Outra característica é que não se nota, visualmente, nenhuma tendência nem ciclo bem definidos.

Realizou-se, a exemplo das demais agências, uma análise estatística das componentes e o resultado está compilado no quadro abaixo.

Quadro 36: Agência 13. Estatística das componentes da série

	Agência 13			
	com outliers		sem outliers	
	Tendência e ciclo	Irregularidade	Tendência e ciclo	Irregularidade
Desvio padrão	33,45	45,88	29,72	36,74
Valor mínimo	-89,59	-242,86	-83,81	-121,77
Valor máximo	288,85	616,40	85,49	123,78

Vê-se, no caso da agência 13, que a componente aleatória (irregularidade) também é muito forte, superando as demais componentes. A seguir têm-se os histogramas da agência.

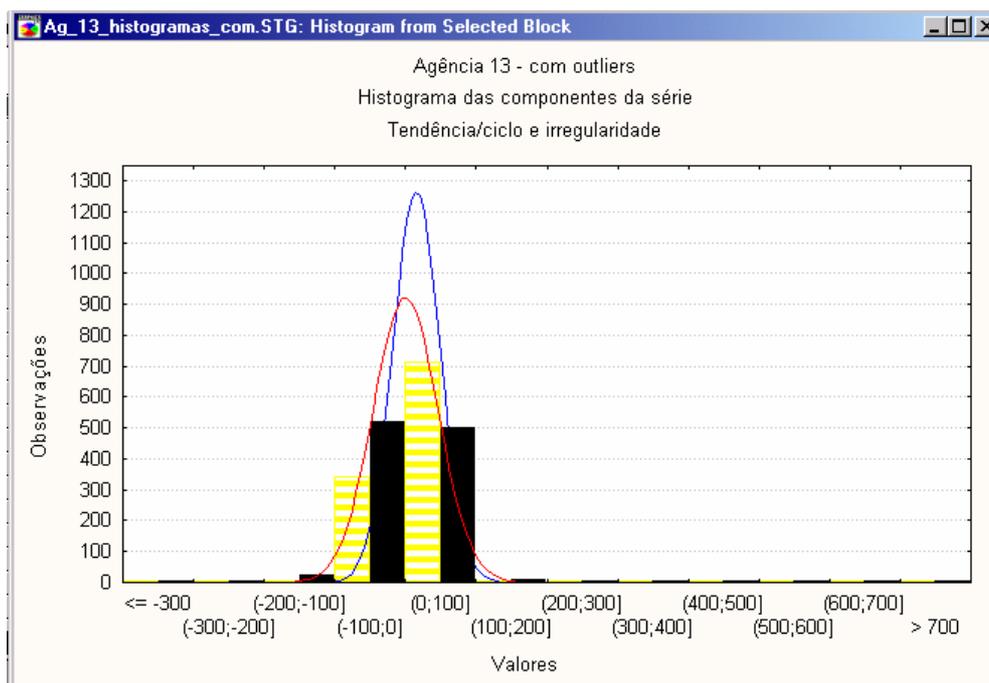


Gráfico 41: Agência 13 (com outliers) – Histograma referente às componentes tendência/ciclo (retângulo tracejado) e irregularidade (retângulo sólido).

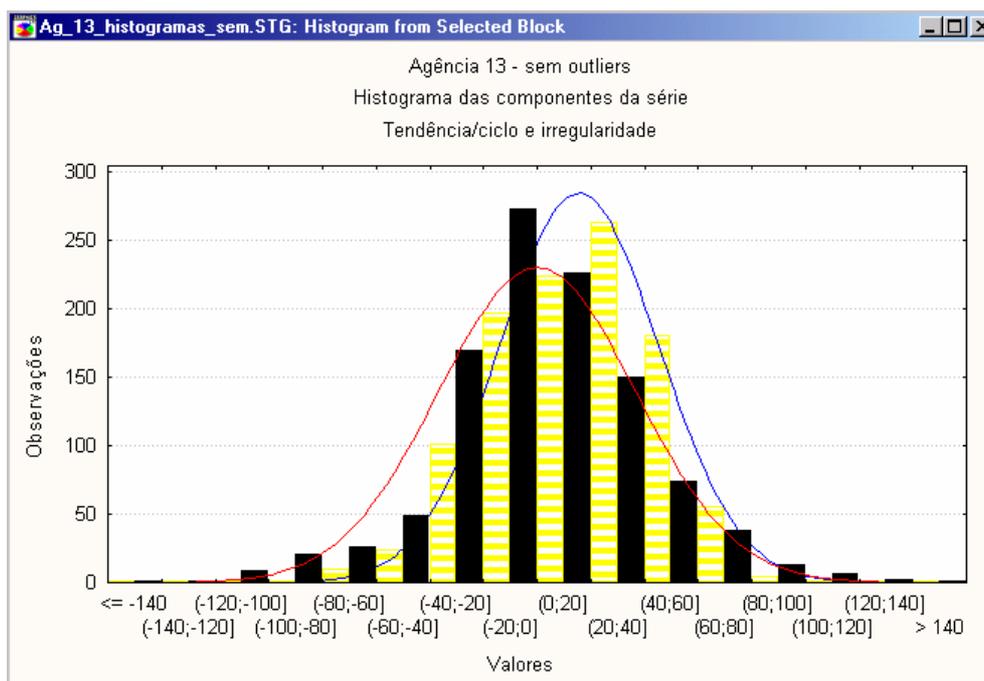


Gráfico 42: Agência 13 (sem outliers) – Histograma referente às componentes tendência/ciclo (retângulo tracejado) e irregularidade (retângulo sólido).

6.3. MODELOS ESTATÍSTICOS UTILIZADOS

Para estudar estatisticamente as séries utilizou-se, como auxílio, o *software forecast.pro*. Trata-se de um aplicativo comercial usado na realização de previsões de séries temporais e dentre as diversas opções disponíveis, a ferramenta executa, automaticamente, a análise dos dados e sugere o melhor modelo a ser aplicado.

A decomposição das séries das agências 02, 06 e 13 realizada pelo *software* apresentou os resultados abaixo:

Quadro 37: Decomposição clássica das séries relativas às agências 02, 06 e 13

		Tendência/Ciclo	Sazonalidade	Irregularidade	Modelo Recomendado
Agência 02	Com outliers	26,52%	1,65%	71,83%	Arima
	Sem outliers	30,63%	1,60%	67,77%	Amortecimento Exponencial
Agência 06	Com outliers	29,46%	0,83%	69,71%	Amortecimento Exponencial
	Sem outliers	26,50%	1,27%	72,23%	Amortecimento Exponencial
Agência 13	Com outliers	19,02%	2,14%	78,84%	Arima
	Sem outliers	22,55%	1,97%	75,48%	Amortecimento Exponencial

Entretanto, no relatório gerado com os resultados das análises, o *forecast.pro* informa que tanto o modelo Arima como o de amortecimento exponencial podem ser aplicados, pois a escolha entre um e outro se deu por diferença muito

pequena. Assim, optou-se pela aplicação de ambos os modelos em todas as séries e os resultados podem ser vistos nos gráficos e quadros adiante. Foram feitas as previsões para os 50 (cinquenta) últimos registros e tratou-se as séries sem *outliers*, ou seja, os valores superiores a 2 desvios padrões foram trazidos para este patamar.

O gráfico abaixo representa os valores reais e previstos, sendo que a linha pontilhada representa o valor real e a linha contínua a previsão.

Gráfico 43: Agência 02 – Gráfico dos dados reais (linha pontilhada) e previstos (linha contínua) (modelo Box e Jenkins).

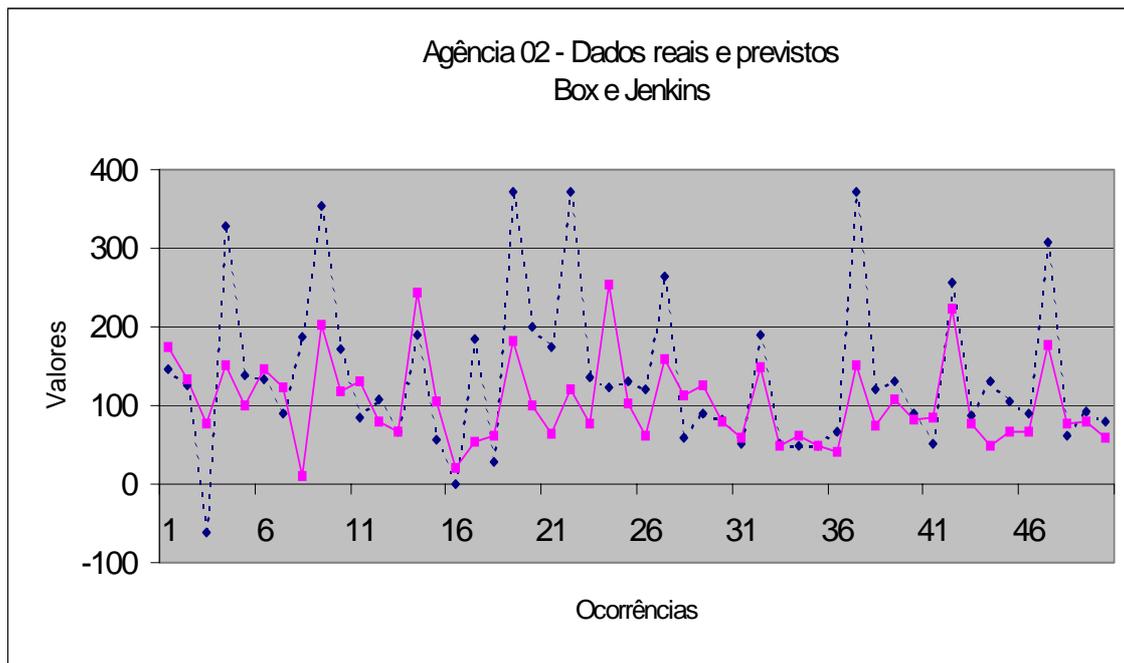


Gráfico 44: Agência 02 – Gráfico dos dados reais (linha pontilhada) e previstos (linha contínua) (modelo amortecimento exponencial).

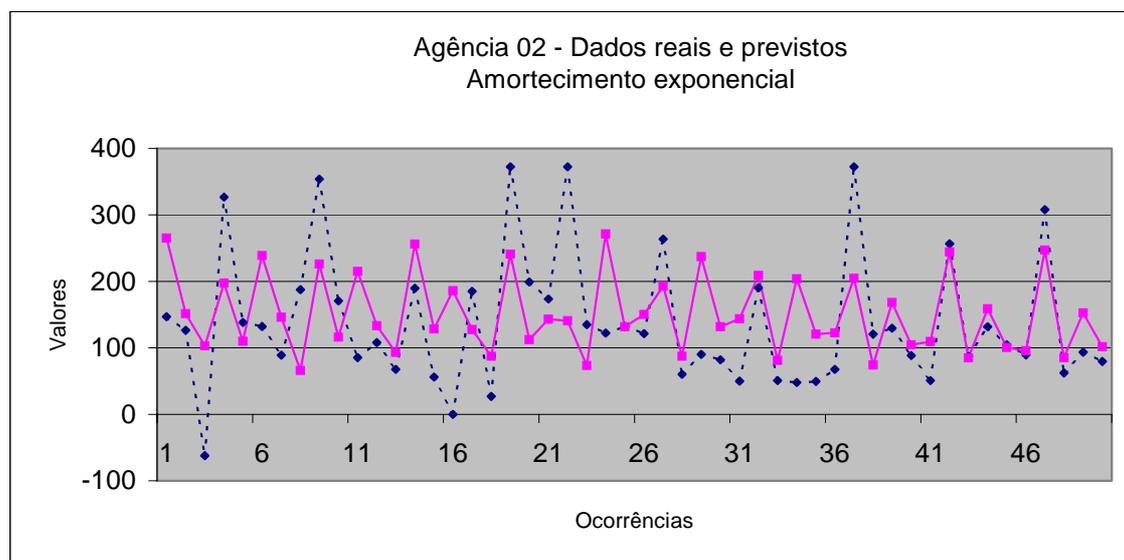


Gráfico 45: Agência 06 – Gráfico dos dados reais (linha pontilhada) e previstos (linha contínua) (modelo Box e Jenkins).

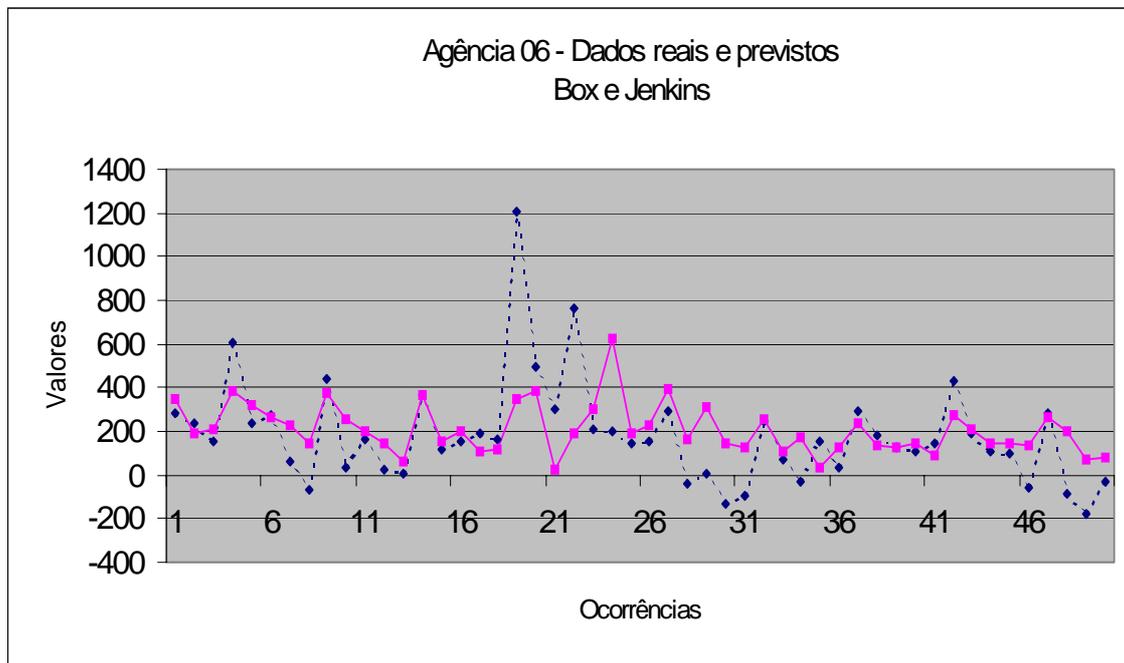


Gráfico 46: Agência 06 – Gráfico dos dados reais (linha pontilhada) e previstos (linha contínua) (modelo amortecimento exponencial)

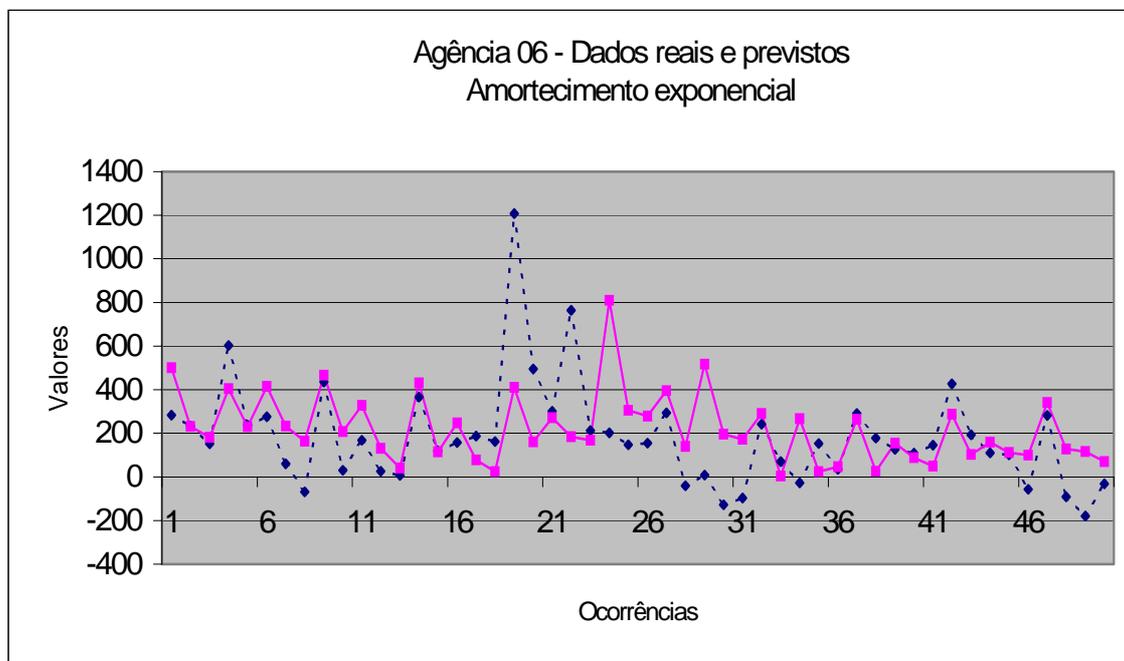


Gráfico 47: Agência 13 – Gráfico dos dados reais (linha pontilhada) e previstos (linha contínua) (modelo Box e Jenkins).

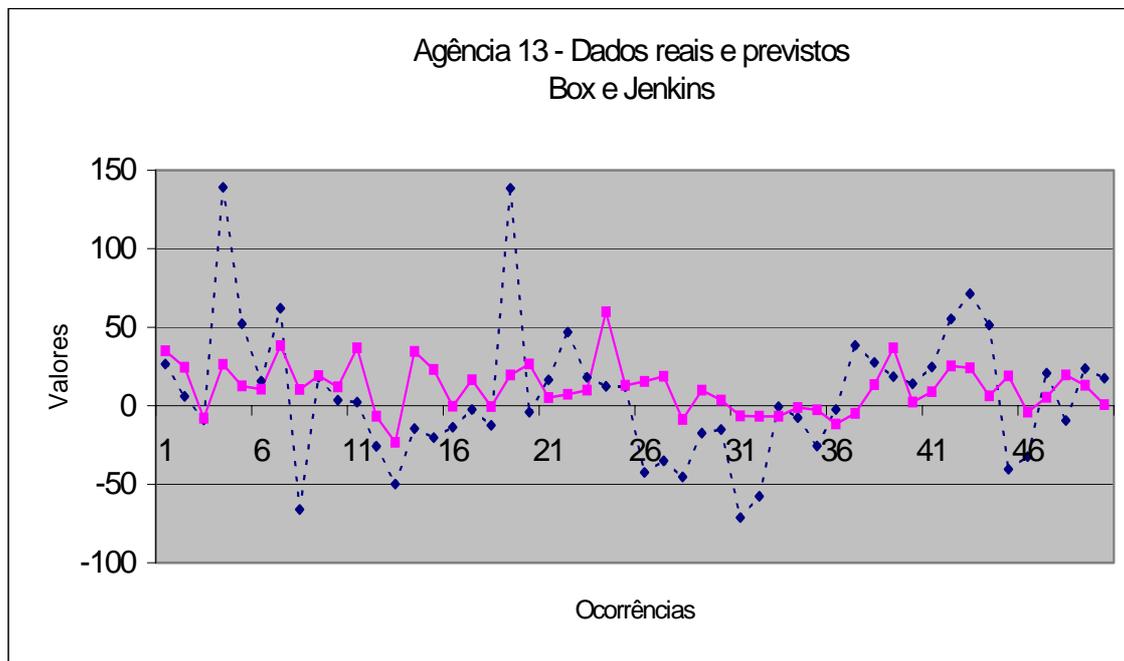
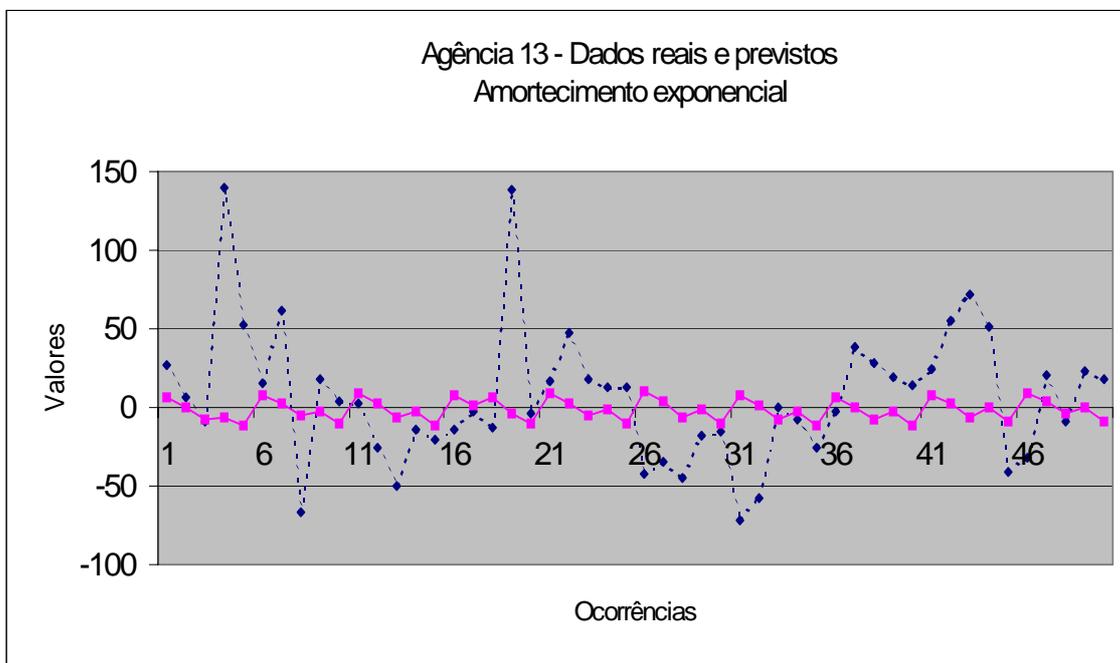


Gráfico 48: Agência 13 – Gráfico dos dados reais (linha pontilhada) e previstos (linha contínua) (modelo amortecimento exponencial).



No item 3, do Anexo, tem-se quadro com os todos os valores previstos pelos modelos. Nas colunas “Acertos” existentes nos quadros está marcado o registro cuja previsão ficou dentro do intervalo de $\pm 20\%$ do valor original.

As previsões utilizando os modelos de Box e Jenkins e amortecimento exponencial não apresentaram resultados satisfatórios. Este fato pode ser atribuído à componente irregularidade, muito marcante nas séries.

6.4. CONCLUSÕES

As séries em estudo possuem particularidades que as diferenciam das séries temporais clássicas, a exemplo de séries de estoques industriais, vendas de produtos, índices financeiros, dados meteorológicos e outras mais. Além da forte presença de irregularidades também existem grandes variações de valores que ocorrem entre um dia e o dia seguinte. Não raro tem-se em d uma posição de milhares de unidades negativas e, em $d + 1$ inverte-se para milhares de unidades positivas, onde d corresponde ao dia vigente. Estes fatores – aleatoriedade e grande oscilação – dificultam o mapeamento da função supostamente geradora da série.

Tendo em vista as características da série, optou-se por se realizar uma análise a respeito da aleatoriedade dos dados, pois na hipótese de se constatar que a série é aleatória a modelagem dos dados se tornaria praticamente impossível.

7. ABORDAGEM NÃO CLÁSSICA DAS SÉRIES EM ESTUDO

Neste capítulo discute-se, inicialmente, os motivos que levaram a escolha dos modelos MLP e Elman para estudo. A seguir, na 2ª seção são apresentados os resultados do treinamento de redes MLP com entradas de dados normalizadas linearmente. No item 3 têm-se os resultados utilizando-se da técnica de CPD-Componente Principal de Discriminação. As seções 4, 5, 6 e 7 apresentam as topologias e resultados obtidos com diversas estratégias de modelagem adotadas. A seção 8 apresenta a abordagem nebulosa utilizada para os dados das agências e no item 9 os respectivos resultados alcançados. Por último, têm-se as conclusões.

7.1. MODELOS E EXPERIMENTOS REALIZADOS

As redes neurais artificiais são classificadas, segundo Simon Haykin (HAYKIN, 1999, p.21-23), em 3 categorias: *single-layer feedforward networks*, *multilayer feedforward networks* e *recurrent networks*. Nesta pesquisa utilizou-se redes do tipo *multilayer feedforward* e *recurrent*.

Não se tem ainda nenhuma fórmula capaz de prever qual a topologia a ser empregada visando à obtenção do melhor resultado para um determinado problema. Muitos estudos são feitos e publicados mostrando algumas soluções para casos específicos, entretanto nenhum pesquisador conseguiu ainda definir parâmetros a serem adotados que resultem em uma rede neural ótima, haja vista a enorme gama de variáveis a serem ajustadas, tais como: quantidade de camadas escondidas, quantidade de neurônios por camada, função de ativação, algoritmo de treinamento, tamanho da “janela” de entrada, método de transformação e normalização de dados e medidas de performance entre outros.

Nesta dissertação alguns modelos de redes foram utilizados e testados, principalmente a MLP – *MultiLayer Perceptron*, da classe de redes *multilayer feedforward networks* e as do tipo “Elman”, categorizadas como *recurrent networks*. Tal escolha se deveu ao fato de serem as mais utilizadas em problemas envolvendo previsão de séries temporais (ZHANG; PATUWO; HU, 1988, p.37) (BRAGA; CARVALHO; LUDEMIR, 1998, p.198-199).

Durante o andamento da pesquisa vários parâmetros foram alterados, sempre visando a melhoria de resultados. Utilizou-se nas redes neurais implementadas, com maior frequência, 2 camadas de neurônios, em razão do fato de redes criadas desta forma possuírem a propriedade de aproximar qualquer função (BRAGA et al *apud* Cybenko, 1998, p. 49) (KÓVÁCS *apud* Kolmogorov, 1996, p.107) (ZHANG; PATUWO; HU *apud* Hornik, 1998, p. 36).

Normalização de dados também foi alvo de especial atenção, haja vista a grande variância existente na série sob estudo. Assim, foram realizados vários testes com diversas normalizações, tanto lineares como não lineares.

Por fim, merece comentário, também, o tratamento dispensado aos *outliers* existentes no conjunto de treinamento. Em alguns casos o procedimento adotado foi simplesmente excluí-los; em outros, não retirá-los mas trazê-los para um patamar pré-definido e, finalmente, em outros, como nos experimentos utilizando-se de lógica nebulosa, trabalhou-se com os dados sem nenhum tratamento para os *outliers*.

7.1.1. Pré-processamento dos Dados

A pesquisa iniciou-se com o tratamento dos dados. As séries estudadas foram entregues em sua forma bruta, sem nenhum cuidado com consistência, checagem de valores etc. Foi necessário, primeiramente, extrair das bases originais os dados de interesse da pesquisa para cada dia da série. Assim, foram separados e compilados os valores relativos aos saldos anteriores, suprimentos, alívios e saldos atuais. Buscou-se consistir estes valores, ou seja, verificar se o saldo final em d-1 correspondia ao saldo inicial em d. Verificou-se, também, se todas as movimentações diárias estavam corretas, ou seja, se a aritmética dos valores (saldo anterior + movimentação do dia = saldo atual) estava correta.

Após esta etapa, houve a necessidade de checagem dos valores considerados, em um primeiro momento, como espúrios. Buscou-se na base de dados todos os registros cuja movimentação e/ou saldo fossem avaliados como discrepantes em relação aos demais. Estes dados foram analisados e checados com os registros originais a fim de se evitar erros de transcrições e/ou escalas.

Foram avaliados, também, os efeitos dos feriados nas séries. Procurou-se detectar alguma relação entre os feriados e as agências, ou seja, se todas se comportam

da mesma maneira para um mesmo feriado. Se um determinado feriado se comporta igualmente em todos os anos etc.

Terminada esta etapa passou-se a estudar qual a melhor transformação que se poderia aplicar nos dados. Foram feitos estudos com diversas formas de normalização, tanto lineares como não lineares. Criou-se funções e avaliou-se as suas distribuições. Objetivou-se, nesta fase, preparar os dados de forma a facilitar o processamento da rede, bem como a obter-se melhores resultados.

Este é um trabalho desgastante e que consome muitas horas de dedicação pois a cada função definida há necessidade de se testar a sua performance. Caso não responda bem, volta-se ao início do problema e novas funções, testes e tratamentos devem ser criados e avaliados.

7.1.2. Redes Neurais Artificiais

Esta pesquisa tratou, basicamente, de 2 modelos neurais: *MultiLayer Perceptron* e redes “Elman”. Ambas são descritas na literatura como eficazes para o tratamento de séries temporais. Outros modelos de rede, tipo SONN (*Self-Organizing Neural Networks*), sugerido em junho de 1989, por Manoel Fernando Tenório e WeiTsih Lee (TENÓRIO, 1989), bem como a rede de Kohonen (SOM – *Self-Organizing Map*) para clusterização dos dados de entrada, também foram testados mas os resultados não se mostraram promissores.

Dentre as categorias de redes mais utilizadas nesta dissertação – MLP e Elman - centenas de protótipos foram criados e testados. Trabalhou-se com redes de 2 e 3 camadas e com vários algoritmos de treinamento. Alterou-se, à exaustão, a quantidade de neurônios, a taxa de aprendizado, as funções de propagação etc.

As redes de retropropagação, onde se inclui a *MLP-MultiLayer Perceptron*, também são chamadas de *feedforward*, pois o processamento da informação dá-se no sentido progressivo, através das interconexões entre os neurônios das camadas adjacentes. Cada unidade de uma camada é conectada para frente a cada unidade da camada seguinte. As ativações fluem da camada de entrada para a camada oculta, e daí para a camada de saída. Como sempre, o conhecimento da rede é codificado nos pesos sobre conexões entre unidades.

Entretanto, apesar de eficiente para classificação, reconhecimento de padrões etc. uma clara deficiência das redes *feedforward* é a dificuldade de fazer com que os modelos lidem com tarefas temporais. Muitos dos algoritmos de treinamento não são capazes de implementar mapeamentos dinâmicos, como, por exemplo, o algoritmo de *backpropagation*, que pode apenas aprender mapeamentos estáticos. Um artifício utilizado para processamento temporal utilizando estas redes envolve o uso de janelas de tempo, onde a entrada da rede utiliza trechos dos dados temporais como se eles formassem um padrão estático (BRAGA; CARVALHO; LUDEMIR, 1998, p. 198). Esta dissertação utilizou-se desta metodologia no trato das redes MLP.

Para lidar com a limitação descrita anteriormente, surgiram as redes recorrentes, ou redes com laços (*loops*). Redes deste tipo foram descritas por Jordan (1988) e Elman (1990) e possuem unidades de contexto, que são elementos de processamento que relembram atividades passadas, simulando um certo tipo de memória de curto prazo. Estas unidades de contexto são importantes quando o processo de aprendizagem dos exemplos ocorre ao longo do tempo.

7.2. EXPERIMENTO 1

Rede MLP com 3 Camadas e Normalização Linear

A principal característica da primeira rede treinada foi o fato de ter-se retirado da série todos os registros superiores a 1 desvio padrão mais R\$ 100.

Buscou-se, também, realizar testes com várias topologias de redes a fim de se analisar as características dos dados e do modelo. Assim, foram geradas 50 redes para cada agência e para cada tipo de normalização. Todas com os pesos iniciais aleatórios.

O processo começou com 10 neurônios na primeira camada, 5 na segunda e 1 na saída, executando-se 5 vezes o treinamento para cada rede. Em seguida acrescentou-se 5 neurônios na segunda camada e repetiu-se o procedimento. Posteriormente colocou-se mais 10 neurônios na primeira camada e voltou-se à segunda camada para 5 neurônios. Assim continuou-se até se chegar ao limite de 50 neurônios na primeira camada e 10 na segunda. Em suma: a 2^a. camada oscilou em 5 e 10 neurônios para cada ciclo de treinamento enquanto a 1^a. camada foi aumentando de 10

neurônios por ciclo. A camada de saída não foi alterada, permanecendo com um único neurônio. Dessa forma, foram treinadas redes com a arquitetura de 10 x 5 x 1, 10 x 10 x 1,..... até o máximo de 50 x 10 x 1.

Na etapa de validação do modelo foi considerado como acerto os valores previstos pela rede que se enquadravam no intervalo de 20% para mais ou para menos do valor real. Trabalhou-se, nesta fase, somente com normalizações lineares cujas fórmulas estão descritas a seguir.

1. $\frac{x_i}{|x_{máx}|}$;
2. $\frac{x - x_{mín}}{x_{máx} - x_{mín}}$;
3. $\left[\left(\frac{x - x_{mín}}{x_{máx} - x_{mín}} \right) - 0.5 \right] * 2$;
4. $\frac{x - \bar{x}}{\sigma}$, onde σ corresponde ao desvio padrão da série.

Adiante têm-se os principais parâmetros utilizados na modelagem das redes. Cabe destacar que todas as redes constantes desta etapa tiveram seus treinamentos interrompidos em função de haverem alcançado o número máximo de épocas definido (2.000) e não por terem atingido o ponto de parada ($goal = 0,001$).

Quadro 38: Principais parâmetros adotados na construção das RNA.

Tipo de Rede	MLP-Feed-Forward Backpropagation		
Tamanho da "janela" de entrada	5 elementos		
Parada (<i>goal</i>)	0,001		
Taxa de aprendizado	0,01		
Épocas	2.000		
Funções de propagação	Tansig, tansig e purelin		
Função de treinamento	RPROP - Resilient Backpropagation		
Quantidade total de registros	811		
	Ag. 02	Ag. 06	Ag. 13
Dados de treinamento	581	550	638
Dados de teste	145	138	159
Dados excluídos (<i>outliers</i>)	85	123	14

No item 4, do Anexo, se encontra o resumo das topologias que apresentaram os melhores resultados bem como o percentual de registros cujos valores situaram-se no intervalo entre -20% e + 20% do valor real.

A fim de verificar a estabilidade das redes treinadas, foram executadas as melhores topologias por mais 5 vezes. Optou-se, nestes novos treinamentos, usar a normalização 3 em todas as agências. Tal escolha deve-se ao fato de cada agência haver tido o seu melhor desempenho com um tipo de normalização distinta, sendo que o segundo melhor resultado para as agências 02 e 06 é justamente aquele cujos dados foram normalizados para o intervalo $[-1, 1]$ – 3ª normalização e a agência 13 já ter esta normalização como a de melhor performance. Outra mudança promovida foi com relação à retirada dos *outliers*. Ao invés de se usar como parâmetro para limitar os valores da base a soma de 1 desvio padrão mais R\$100, excluiu-se os registros cujos valores eram superiores a 2 desvios padrões. Cabe comentar que os dados para trabalho estão em milhares de reais, logo, quando se refere ao limite como sendo de 1 DP mais R\$100 na realidade o parâmetro fixo significa 100 mil reais.

Por fim, verifica-se que a agência 06 possui duas topologias que obtiveram o mesmo índice de acerto. Neste caso optou-se por treinar a de menor número de neurônios. Adiante tem-se um quadro mostrando os principais parâmetros adotados.

Quadro 39: Principais parâmetros adotados na construção das RNA.

Tipo de Rede	MLP-Feed-Forward Backpropagation		
Tamanho da "janela" de entrada	5 elementos		
Parada (<i>goal</i>)	0,001		
Taxa de aprendizado	0,01		
Épocas	2.000		
Funções de propagação	Tansig, tansig e purelin		
Função de treinamento	RPROP - Resilient Backpropagation		
Quantidade total de registros	811		
	Ag. 02	Ag. 06	Ag. 13
Dados de treinamento	615	615	615
Dados de teste	154	154	156
Dados excluídos (<i>outliers</i>)	42	42	40
Topologia	10 x 5 x 1	10 x 10 x 1	10 x 5 x 1

Os resultados alcançados nestes testes estão mostrados no item 4, do Anexo, e, apesar de em nenhum momento a rede haver convergido para o índice de parada (*goal*) definido, os modelos apresentaram boa performance durante a fase de treinamento, com índices de acerto superiores a 90%.

O fato de os índices de performance do treinamento terem sido bons significa que a rede conseguiu aproximar bem a função. Entretanto, quando do teste, os resultados se mostraram aquém das expectativas, mostrando que a rede não estava apresentando uma boa generalização. Cabe lembrar que a baixa generalização não

decorre da existência de poucos neurônios – caso mais comum, pois foram testadas as mesmas redes com a quantidade de 1 a 50 neurônios. Atribui-se tal característica a grande oscilação existente nas séries sob estudo.

Em vista das performances haverem sido extremamente insatisfatórias, procurou-se verificar se a direção apontada na previsão referente ao melhor resultado de cada agência correspondia àquela referente aos dados reais, ou seja, se a diferença entre o valor corrente e o anterior (d_t e d_{t-1}) era negativa ou positiva em ambos os conjuntos, o de teste e o previsto. Obteve-se o resultado descrito no item 4, do Anexo.

Vê-se, no quadro acima, que mesmo a previsão unicamente dos sinais dos valores das séries não foram bem mapeados pela rede. Tal problema pode ser atribuído a característica das séries de possuírem grande oscilação de valores, que se alteram de forma abrupta, ou seja, em d tem-se valores negativos para, logo em seguida, voltar para positivo e, alguns poucos dias depois, tornarem-se negativo novamente. Não há um crescimento e decrescimento gradual dos valores. Tratar estas inflexões é muito difícil e a rede demonstrou isto claramente.

7.3. EXPERIMENTO 2

Componente Principal de Discriminação – Janela de 5 elementos

Nesta etapa do trabalho, optou-se por se verificar qual a quantidade ideal de neurônios para que a rede apresentasse os melhores resultados. Para resolver este problema escolheu-se desenvolver redes neurais usando CPD (Componente Principal de Discriminação). Dessa forma, foram criadas 50 redes para cada agência, com até cinquenta neurônios que foram sendo inseridos um a um após o término de cada treinamento, ou seja, após treinar a rede calcula-se os índices de acerto, acrescenta-se um neurônio e inicia-se novo treinamento. A fim de melhor exemplificar o algoritmo, segue abaixo os passos adotados (THOMÉ, 2001, p. 1).

8. Coloque 1 único neurônio na camada escondida;
9. Treine a rede até que o erro estabilize;
10. Calcule o desempenho no teste;
11. Salve os pesos da rede, a evolução do erro no treinamento e o desempenho no teste;

12. Fixe os pesos da primeira camada que incidem sobre o neurônio da camada escondida;
13. Verifique se o desempenho obtido melhorou ou está estabilizado;
14. Se melhorou, acrescente mais um neurônio e volte ao passo 2;
15. Se piorou, pare.

Os principais parâmetros da sua arquitetura estão descritos abaixo.

Quadro 40: Principais parâmetros na criação de redes para verificar a quantidade ideal de neurônios. Janela de 5.

Tipo de Rede	MLP-Feed-Forward Backpropagation		
Tamanho da "janela" de entrada	5 elementos		
Parada (goal)	0,001		
Taxa de aprendizado	0,01		
Épocas	2.000		
Funções de propagação	Tansig e purelin		
Função de treinamento	RPROP - Resilient Backpropagation		
Normalização dos dados	$\left[\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}} \right) - 0.5 \right] * 2$		
Quantidade total de registros	811		
	Ag. 02	Ag. 06	Ag. 13
Dados de treinamento	615	615	615
Dados de teste	154	154	156
Dados excluídos (outliers)	42	42	40

No item 5, do Anexo, têm-se os resultados da agência 02, 06 e 13 que tiveram como média de acertos os valores de 8,66, 11,10 e 13,70 e desvios padrão de 3,02, 3,73 e 2,25.

Os resultados obtidos não se mostraram razoáveis e, em função disto, escolheu-se alterar o tamanho da “janela” de entrada. Assim, novos testes idênticos ao anterior foram realizados, sendo que foram utilizados 10 elementos de entrada.

7.4. EXPERIMENTO 3

Componente Principal de Discriminação – Janela de 10 elementos

O uso de “janelas” de 10 elementos se deve ao fato deste tamanho corresponder a períodos múltiplos semanais (dias úteis) e, acredita-se, que possa existir alguma similaridade entre os períodos.

Abaixo se têm os principais parâmetros utilizados.

Quadro 41: Principais parâmetros na criação de redes para verificar a quantidade ideal de neurônios. Janela de 10.

Tipo de Rede	MLP-Feed-Forward Backpropagation		
Tamanho da "janela" de entrada	10 elementos		
Parada (goal)	0,001		
Taxa de aprendizado	0,01		
Épocas	2.000		
Funções de propagação	Tansig e purelin		
Função de treinamento	RPROP - Resilient Backpropagation		
Normalização dos dados	$\left[\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}} \right) - 0.5 \right] * 2$		
Quantidade total de registros	811		
	Ag. 02	Ag. 06	Ag. 13
Dados de treinamento	615	615	615
Dados de teste	154	154	156
Dados excluídos (outliers)	42	42	40

A agência 02 abaixo teve como média de acertos o valor de 15,48 e desvio padrão de 3,37.

No item 6, do Anexo, têm-se os resultados da agência 02, 06 e 13 que tiveram como média de acertos os valores de 15,48, 12,50 e 13,44 e desvios padrão de 3,37, 2,23 e 2,25.

O uso da técnica de CPD não se mostrou adequado ao problema. Foram testadas redes com 1 à 50 neurônios na camada de entrada e seria razoável supor encontrar algum intervalo onde os resultados fossem melhores. Entretanto, verificando os quadros de acertos vê-se que a exceção de alguns picos não muito significativos a maioria dos resultados manteve-se constante com qualquer quantidade de neurônios.

7.5. EXPERIMENTO 4

Conjunto de Treinamento com Início de Mês Coincidente

Conforme discutido anteriormente, uma das hipóteses levantadas diz respeito ao fato de existir forte correlação entre os períodos de início de mês assim como entre os dias da semana (segunda-feira, terça-feira, etc).

Baseado nas premissas acima foram desenvolvidas redes com o conjunto de treinamento separado por dia da semana e início de mês, ou seja, foram agrupados todos os meses cujo primeiro dia útil é segunda-feira, outro conjunto onde o primeiro dia útil do mês é uma terça-feira, e assim sucessivamente. Abaixo se tem o quadro mostrando

os meses/anos cujos inícios são coincidentes.

Quadro 42: Meses/anos cujos primeiros dias úteis do mês são coincidentes.

Meses/anos cujo 1º. dia útil foi segunda-feira.							
fev/98	mar/98	mai/98	jun/98	ago/98	jan/99	fev/99	mar/99
abr/99	mai/99	ago/99	nov/99	jan/00	abr/00	jul/00	out/00
Meses/anos cujo 1º. dia útil foi terça-feira							
set/98	nov/98	dez/98	jun/99	fev/00	mai/00	ago/00	jan/01
Meses/anos cujo 1º. dia útil foi quarta-feira							
abr/98	jul/98	set/99	dez/99	mar/00	nov/00		
Meses/anos cujo 1º. dia útil foi quinta-feira							
out/98	jul/99	jun/00	jan/00	mar/01			
Meses/anos cujo 1º. dia útil foi sexta-feira							
jan/98	out/99	set/00	dez/00				

Em função do reduzido conjunto de dados que se teria para treinar uma rede referente às terças-feiras, quartas-feiras, quintas-feiras e sextas-feiras, optou-se por realizar, num primeiro momento, somente o treinamento para os meses cujo início ocorreu numa segunda-feira. Caso se obtivesse resultados promissores seria avaliado o tratamento a ser dado para os demais dias da semana. Poderia-se, inclusive, juntar dois ou mais dias para efeito de aumentar o conjunto e, então, adicionar uma entrada para auxiliar a rede a identificar os dias separadamente no conjunto.

Outro ponto que merece destaque diz respeito ao tratamento dos *outliers*. A fim de não diminuir ainda mais a massa de treinamento optou-se por considerar todos os registros maiores que duas vezes o desvio padrão como tendo este valor, ou seja, os dados acima deste patamar foram considerados como sendo iguais a duas vezes o desvio padrão do conjunto da agência.

A mesma estratégia de treinamento adotada anteriormente foi repetida neste experimento, ou seja, treinou-se redes de 1 a 50 neurônios na primeira camada. Considerou-se, também, como acerto na previsão, os valores que se situavam no intervalo de -20% a $+20\%$ do real.

Abaixo tem-se o quadro com as principais características das redes implementadas.

Quadro 43: Principais parâmetros na criação de redes com conjunto de treinamento somente dos meses cujo 1º. dia útil seja a segunda-feira.

Tipo de Rede	MLP-Feed-Forward Backpropagation
Tamanho da "janela" de entrada	10 elementos
Parada (goal)	0,001
Taxa de aprendizado	0,01
Épocas	2.000
Funções de ativação	Tansig e purelin
Função de treinamento	RPROP - Resilient Backpropagation
Normalização dos dados	$\left[\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}} \right) - 0.5 \right] * 2$
Quantidade total de registros	811
Dados de treinamento	252
Dados de teste	55

Os resultados obtidos não se mostraram promissores pois a agência 02 teve um índice de acerto de 14,6% dos registros, correspondente à previsão correta de apenas 8 registros, em um total de 55. A rede que obteve melhor performance foi constituída por 27 neurônios na primeira camada e 1 neurônio na 2ª camada.

A agência 06 foi a que melhor respondeu aos treinamentos obtendo um acerto de 20,0% dos dados previstos. Este percentual corresponde a 11 registros com duas configurações que obtiveram a mesma performance: a 1ª. foi de 2 x 1 e a 2ª. de 4 x 1 neurônios.

Por fim, tem-se a agência 13 com 18,2% de acertos – 10 registros corretos. Duas foram as configurações que chegaram a este patamar, a rede de 10 x 1 e a de 48 x 1.

O fraco desempenho obtido com esta estratégia desestimulou o prosseguimento com os demais registros.

7.6. EXPERIMENTO 5

Redes "Elman" – Normalização Linear

Em vista do baixo desempenho obtido até o momento, pesquisou-se novos modelos de redes para serem aplicados ao problema. Escolheu-se, dentre as redes do tipo recorrente, o modelo Elman.

Neste treinamento usou-se novamente a premissa de substituir os valores superiores a 2 desvios padrões por este limite. Foram treinadas 50 redes iniciando-se a

primeira camada com apenas um neurônio e a cada novo treinamento foi-se adicionando um outro neurônio à rede. Na camada de contexto foi inserido 1 neurônio. Os valores previstos foram considerados corretos se estivessem no intervalo de -20% a 20% do real. As demais características básicas da rede estão apresentadas a seguir.

Quadro 44: Principais parâmetros na criação de redes "Elman".

Tipo de Rede	Elman
Tamanho da "janela" de entrada	10 elementos
Parada (goal)	0,001
Taxa de aprendizado	0,01
Épocas	2.000
Funções de ativação	Tansig e purelin
Função de treinamento	Gradiente descendente
Normalização dos dados	$\left[\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}} \right) - 0.5 \right] * 2$
Quantidade total de registros	811
Dados de treinamento	649
Dados de teste	162

A agência 02 abaixo teve como média de acertos o valor de 15,40 e desvio padrão de 3,61.

No item 7, do Anexo, têm-se os resultados da agência 02, 06 e 13 que tiveram como média de acertos os valores de 15,40, 15,66 e 13,16 e desvios padrão de 3,61, 2,65 e 2,64.

Procurou-se, novamente nestes testes, verificar se a direção apontada na previsão referente ao melhor resultado de cada agência correspondia àquela referente aos dados reais, ou seja, se a diferença entre d_t e d_{t-1} era negativa ou positiva em ambos os conjuntos, o de teste e o previsto. Obteve-se o seguinte resultado:

Quadro 45: Acertos na direção dos dados previstos pela melhor rede Elman para d_t e d_{t-1} .

	Acertos (Direção)	% Acertos	Erros (Direção)	% Erros
Ag. 02	57	37,50	94	62,50
Ag. 06	53	34,87	98	65,13
Ag. 13	75	49,34	76	50,66

A normalização linear adotada fez com que os dados se concentrassem próximos de zero. Com isto, a rede tendia a gerar uma curva de amplitude muito baixa, não tratando os valores que estavam mais longe. Acredita-se que um dos motivos da baixa performance seja em função do tipo de normalização adotada. Assim, promoveu-

se alguns testes com os dados normalizados através de equações não lineares visando comprovar esta hipótese.

7.7. EXPERIMENTO 6

Redes MLP e "Elman" – Normalização Não Linear

Nesta seção serão apresentados os resultados obtidos através do treinamento de redes MLP e Elman com o conjunto de treinamento normalizado por funções não lineares. Tais normalizações encontram-se descritas abaixo.

1. $\sqrt{\text{abs}(\text{reg})}$;
2. $\exp(-(\text{abs}(\text{reg} + \text{abs}(\text{média})) / \text{variância}))$;
3. $\exp(-((\text{reg} + \text{abs}(\text{média})) ** 2 / \text{variância}))$;
4. $\exp(-(\text{abs}(\text{reg} + \text{abs}(\text{média})) / \text{desvio padrão}))$.

Adotou-se para tratamento dos *outliers* trazer estes valores para o patamar de 2 desvios padrões. A verificação da performance das redes foi obtida considerando como corretos os valores previstos que se situavam no intervalo de -20% a + 20%. Por fim, estas redes foram construídas inicialmente com 1 único neurônio na primeira camada e, a cada treinamento concluído, iniciava-se outro com a adição de mais um neurônio na primeira camada, até o máximo de 50 neurônios. Para o caso das redes Elman, a camada de contexto possuía 1 neurônio.

Abaixo se encontra o resumo dos demais parâmetros adotados.

Quadro 46: Principais parâmetros na criação de redes de MLP e Elman com uso da fórmula de normalização 1.

Tipos de Redes	MLP e Elman
Tamanho da "janela" de entrada	10 elementos
Parada (<i>goal</i>)	0,001
Taxa de aprendizado	0,01
Épocas	2.000
Funções de ativação	Tansig e logsig
Função de treinamento	Rprop (MLP) e Gradiente descendente (Elman)
Normalização dos dados	Fórmulas 1, 2, 3 e 4 descritas acima
Quantidade total de registros	811
Dados de treinamento	649
Dados de teste	162

Verificou-se que a performance obtida melhorou, em alguns casos, com esta nova abordagem – normalização não linear. Entretanto, conforme pode ser visto, os resultados ainda estão muito aquém do desejado, sendo que a agência 02 teve como pico acertos de 25,7%; a agência 06 teve o seu melhor resultado de 30,3% e a agência 13 de 21,1%. Chamou a atenção, também, o fato de as agências não terem se comportado homoganeamente com relação ao tipo de normalização adotada. Enquanto as agências 02 e 06 tiveram os melhores resultados com a normalização 3 a agência 13 apresentou melhor performance com a normalização 1.

Contrariando as expectativas, o modelo “Elman” não foi superior no caso da agência 13, tendo a rede MLP obtido melhor performance. Para as demais agências apesar de a performance do modelo Elman ter sido superior, a diferença entre ele e o modelo MLP foi insignificante.

Resumo das performances alcançadas está compilado no item 8, do Anexo. Os números constantes da coluna “Norma” (1ª coluna) correspondem às fórmulas de normalização descritas no início da seção.

Complementando os estudos promovidos acima, procurou-se ratificar os melhores resultados e, para tanto, treinou-se as mesmas redes que obtiveram as melhores performances.

Escolheu-se treinar 5 redes consecutivas pertencentes ao grupo em que houve a melhor performance. Há de se notar que em função desta escolha nem sempre o intervalo cujas redes foram treinadas nesta etapa corresponde àquele que, individualmente, alcançou o melhor resultado absoluto, ou seja, há casos em que na primeira fase a melhor performance foi obtida com x neurônios mas em razão da existência de resultados mais homogêneos em outro intervalo optou-se por trabalhar com este último conjunto de redes.

Assim, por exemplo, a agência 02 teve o seu melhor resultado com a rede “Elman”, dados normalizados pela fórmula 1 e com 38 neurônios na primeira camada. Entretanto, verificando a quantidade de acertos das redes antecedentes e precedentes vê-se que a performance não foi regular. Optou-se, então, em função disto, por trabalhar com o intervalo de redes cuja quantidade de neurônios na primeira camada variava de 26 a 30 neurônios pois, neste caso, tem-se situações mais regulares. O mesmo tratamento foi feito com a agência 13 e redes “Elman”, com dados normalizados pelas

fórmulas 3 e 4. Os demais corresponderam a intervalos que abrangem a rede que obteve a melhor performance.

Alterou-se o número de épocas para 3.000 ao invés das 2.000 iniciais tendo em vista que as paradas ocorridas foram em função de haver sido atingido o número de épocas e não por haver alcançado a meta estipulada (0,001). Os demais parâmetros estão compilados no quadro a seguir.

Quadro 47: Principais parâmetros das melhores redes de MLP e Elman usadas na primeira fase.

Tipos de Redes	MLP e Elman
Tamanho da "janela" de entrada	10 elementos
Parada (<i>goal</i>)	0,001
Taxa de aprendizado	0,01
Épocas	3.000
Funções de ativação	Tansig, e purelin
Função de treinamento	Rprop (MLP) e Gradiente descendente (Elman)
Quantidade total de registros	811
Dados de treinamento	649
Dados de teste	162

Resultados relativos aos treinamentos realizados com a agência 02, 06 e 13 encontram-se no item 8, do Anexo.

Vê-se, claramente, que estas últimas redes treinadas não mantiveram a robustez. Os resultados obtidos para ratificar a performance dos treinamentos iniciais foram, em sua maioria, inferiores àqueles alcançados na primeira fase.

Acredita-se que os acertos verificados sejam oriundos de coincidência, ou seja, apesar de a rede conseguir mapear o passado com um bom grau de acertividade (boa aproximação da função característica) ela não está bem parametrizada a ponto de prever eficazmente os valores futuros. Entretanto, como a rede sempre dá uma resposta aos dados de entrada, o valor de saída pode estar dentro do intervalo considerado para acerto.

7.8. EXPERIMENTO 7

Redes MLP e Elman – Funções de propagação não lineares

Avaliando os resultados percebeu-se que uma saída linear – purelin – talvez não fosse adequada para servir como função de propagação, pois os valores extremos poderiam não ser representados. Assim, optou-se pelo uso de funções não lineares para

a segunda camada da rede e novos testes foram realizados usando a função *logsig* na segunda camada. Adiante se tem o resumo dos demais parâmetros adotados.

Quadro 48: Principais parâmetros das redes de MLP e Elman com função de propagação não linear na camada de saída.

Tipos de Redes	MLP e Elman
Tamanho da "janela" de entrada	10 elementos
Parada (<i>goal</i>)	0,001
Taxa de aprendizado	0,01
Épocas	2.000
Funções de ativação	Tansig, e logsig
Função de treinamento	Rprop (MLP) e Gradiente descendente (Elman)
Quantidade total de registros	811
Dados de treinamento	649
Dados de teste	162

Cabe destacar que se usou os mesmos conceitos adotados nos treinamentos anteriores, ou seja, executou-se redes com 1 a 50 neurônios para cada tipo de normalização. Foram treinadas redes MLP e Elman e o número de épocas no treinamento se restringiu a 2.000, sendo que em todos os casos a parada se deu em função de haver sido atingido este limite. Os valores da base de dados superiores a 2 desvios padrões foram trazidos para este patamar.

Os quadros com o resumo das melhores redes, para as agências 02, 06 e 13, estão apresentados no item 9, do Anexo.

Os resultados obtidos nestes experimentos também não foram satisfatórios. Novamente, os índices de acertos da rede MLP ficaram muito próximos daqueles alcançados pela rede Elman, demonstrando que uma arquitetura típica para tratamento de séries temporais também não está respondendo bem aos dados.

Cabe comentar que os dados normalizados pela fórmula 1 não conseguiram acertar nenhuma previsão nas 3 agências. Isto demonstra que uma normalização não adequada acarreta um baixo índice de acertos. Isto, inclusive, pode ser o motivo dos baixos índices de performance.

7.9. EXPERIMENTO 8

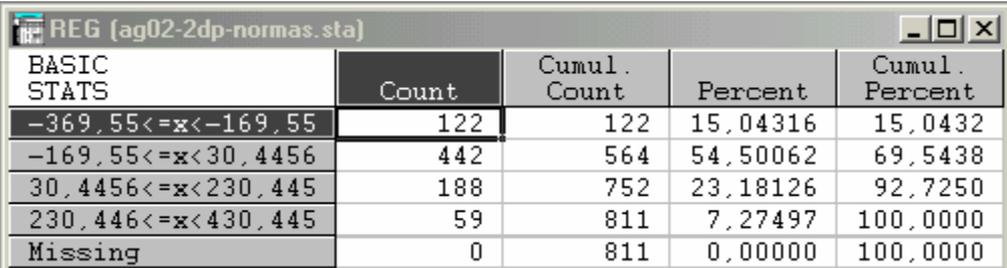
Redes MLP – Verificação da performance da rede através de intervalos

A verificação da performance da rede utilizando-se como parâmetro de acerto o intervalo de $\pm 20\%$ não tem apresentado bons resultados, ou seja, para valores pequenos os limites para se considerar a previsão correta são muito reduzidos.

Como parte dos pontos são de valores baixos, próximos de zero, o intervalo de $\pm 20\%$ é muito pequeno e, às vezes, alguns pontos são considerados como errados e estão bem pertos do valor real. Por exemplo: em 5 de março de 2003 a agência 06 teve um movimento de R\$ 70,68 mil. Neste caso se consideraria correta uma previsão que se situasse no intervalo entre R\$ 56,54 mil e R\$ 84,82 mil.

Ora, esta agência possui movimento máximo de R\$ 4.227,61 mil e média de R\$ 237,05 mil para os 1.059 registros e uma previsão de, por exemplo, R\$ 90 mil seria passível de ser considerada certa. No entanto, para efeito de cálculo da performance, este valor é computado como errado.

Baseado nesta premissa, foram desenvolvidos experimentos onde se dividiu a saída em diversos intervalos (4, 5, 6 e 8) e verificou-se se o valor previsto, após a desnormalização, ficava no intervalo correto. Ou seja, de posse do maior e menor valor da série, calculou-se, no primeiro caso, 4 intervalos de tamanho idêntico e verificou-se se o valor previsto encontrava-se dentro dos limites corretos. Para melhor exemplificar o conceito, vide figura abaixo.



BASIC STATS	Count	Cumul. Count	Percent	Cumul. Percent
$-369,55 <= x < -169,55$	122	122	15,04316	15,0432
$-169,55 <= x < 30,4456$	442	564	54,50062	69,5438
$30,4456 <= x < 230,445$	188	752	23,18126	92,7250
$230,446 <= x < 430,445$	59	811	7,27497	100,0000
Missing	0	811	0,00000	100,0000

Figura 15: Agência 02 – Separação dos dados em 4 intervalos.

Cabe ressaltar que foram construídas redes com 1 a 20 neurônios na camada escondida. Adiante se tem o resumo dos parâmetros da rede adotados.

Quadro 49: Principais parâmetros das redes de MLP com cálculo de performance através de intervalos.

Tipos de Redes	MLP
Tamanho da "janela" de entrada	5 elementos
Parada (goal)	0,001
Taxa de aprendizado	0,01
Épocas	3.000
Funções de ativação	Tansig e purelin
Função de treinamento	Rprop
Normalização dos dados	$\exp(-(\text{abs}(\text{reg} + \text{abs}(\text{média}))/\text{desvio padrão}))$
Quantidade total de registros	811
Dados de treinamento	649
Dados de teste	162

Os resultados das agências 02, 06 e 13 para 4, 5, 6 e 8 intervalos estão compilados no item 10, do Anexo.

Os índices de acertos, adotando esta metodologia, foram superiores, conforme já se esperava, pois o intervalo para se considerar a previsão como correta foi aumentado significativamente. Entretanto, há necessidade de se aumentar o número de intervalos para melhor ajuste dos valores, pois trabalhar-se com 4, 5, 6 ou 8 intervalos apenas faz com que os valores previstos possam vir a ficar muito divergentes do valor real. Por exemplo: a divisão em 8 intervalos para a agência 6 significa que cada um deles será de R\$ 300 mil, o que é considerado um valor muito alto para uso. Por outro lado, se reduzirmos estes limites, aumentando o número de intervalos, a performance cai. Vê-se que, para esta mesma agência 6, a divisão dos valores em 6 intervalos acarretou um acerto máximo da ordem de 62,42% e, quando se aumentou o número de intervalos para 8, o índice caiu para 49,68%.

7.10. EXPERIMENTO 9

Redes MLP com quantização dos dados

Objetiva-se neste experimento definir a melhor quantidade de intervalos para classificação dos dados. Para tanto, treinou-se redes com 1 a 20 neurônios na camada escondida e escolheu-se a arquitetura que melhor performance apresentou para os dados, considerando como certo o intervalo de $\pm 20\%$. De posse dos dados de saída da melhor rede fez-se uma classificação com a divisão dos dados entre 2 e 10 intervalos.

Os parâmetros da rede utilizados neste experimento estão descritos abaixo.

Quadro 50: Principais parâmetros das redes de MLP com cálculo de performance através da divisão de 2 a 10 intervalos .

Tipos de Redes	MLP
Tamanho da "janela" de entrada	5 elementos
Parada (<i>goal</i>)	0,001
Taxa de aprendizado	0,01
Épocas	3.000
Funções de ativação	Tansig e purelin
Função de treinamento	Rprop
Normalização dos dados	$\exp(-(\text{abs}(\text{reg} + \text{abs}(\text{média}))/\text{desvio padrão}))$
Quantidade total de registros	811
Dados de treinamento	649
Dados de teste	162

No item 11, do Anexo, têm-se os resultados alcançados e, como era de se esperar, o aumento da quantidade de intervalos reduziu o índice de acertos. As melhores performances ocorreram com, no máximo, 3 intervalos o que torna inviável a sua utilização pois os valores contemplados no intervalo ficam muito grandes. Para poder dar um tratamento baseado em intervalos tem-se que trabalhar com, no mínimo, 10 intervalos de valores e, neste caso, a performance não satisfaz.

7.11. EXPERIMENTO 10

Redes MLP e "Elman" com binarização

A estratégia adotada neste experimento foi a de binarizar os dados de entrada, após a sua normalização. Ou seja, criou-se 8 intervalos e classificou-se cada registro dentro de um intervalo correspondente aos algarismos binários (000, 001, 010,....., 111). Aplicou-se, então, estes valores nos modelos neurais.

As saídas da rede (3 no total) também foram binarizadas para 0 se menor ou igual a 0,5 e 1 para os demais valores. Aqui também se trabalhou com 1 a 20 neurônios.

Os parâmetros da rede utilizados neste experimento estão descritos abaixo.

Quadro 51: Principais parâmetros das redes de MLP e Elman no experimento 10.

Tipos de Redes	MLP e Elman
Tamanho da "janela" de entrada	5 elementos
Parada (<i>goal</i>)	0,001
Taxa de aprendizado	0,01
Épocas	3.000
Funções de ativação	Tansig e purelin
Função de treinamento	Rprop
Normalização dos dados	$\log_{10}(\text{reg} + \text{abs}(\min(\text{reg})) + 1)$
Quantidade total de registros	811
Dados de treinamento	649
Dados de teste	162

No item 12, do Anexo, têm-se os resultados alcançados para as agências 02, 06 e 13.

A exceção da agência 13, vê-se que os resultados apresentaram uma performance razoável para a divisão dos dados em 8 intervalos. Entretanto, ao aumentar-se o número de intervalos os índices caem drasticamente. Ainda, assim, optou-se por se testar outras formas de normalização, a fim de verificar se se consegue melhoria nos índices de acertos.

7.12. EXPERIMENTO 11

Redes MLP com binarização

Este experimento é similar ao anterior, entretanto aqui utilizou-se somente modelos MLP pois a diferença de performance com os modelos “Elman” foi insignificante. Trabalhou-se, também, com outros tipos de normalizações e aumentou-se o número de épocas durante a fase de treinamento.

Por fim, alterou-se, também, a quantidade de neurônios, trabalhando-se com 1 a 30 na entrada e 4 na saída da rede, sendo que, agora, cada saída terá apenas um valor 1 (o maior valor entre os 4 neurônios) e os demais 0, ou seja, reduziu-se o intervalo de saída para 4.

Os parâmetros da rede utilizados neste experimento estão descritos abaixo.

Quadro 52: Principais parâmetros das redes de MLP no experimento 11.

Tipos de Redes	MLP – Multilayer Perceptron
Tamanho da "janela" de entrada	5 elementos
Parada ("goal")	0,001
Taxa de aprendizado	0,01
Épocas	5.000
Funções de ativação	Logsig e logsig
Função de treinamento	RPROP - Resilient Backpropagation
Normalização dos dados	Z-Score (X / abs(Xmax)) (X - Xmin) / (Xmax - Xmin) Log10(reg + abs(min(reg)) + 1) sqrt(abs(reg))
Quantidade total de registros	811
Dados de treinamento	649
Dados de teste	162

No item 13, do Anexo, têm-se os resultados alcançados para as agências 02, 06 e 13.

Os resultados também não se mostraram promissores utilizando esta estratégia. Os índices estão baixos se considerar que a rede foi treinada para classificar os registros em apenas 4 intervalos.

Foram testadas estas mesmas redes para janelas de 10, 15 e 20 elementos mas os resultados não se mostraram bons, por este motivo não serão apresentados.

7.13. EXPERIMENTO 12

Redes MLP com média móvel

Tendo em vista a grande variância dos dados, buscou-se neste experimento trabalhar com médias móveis. Para isto, o conjunto de treinamento foi normalizado e modificado para se trabalhar com médias móveis de 5, 10, 15 e 20 dias, além de se ter treinado a rede sem média móvel, só com a normalização. Além disso, realizou-se 10 treinamentos para cada tipo de rede a fim de se avaliar a robustez dos resultados. Por último, trabalhou-se com 3 camadas de neurônios formando uma arquitetura do tipo 20 x 10 x 1.

Os parâmetros da rede utilizados neste experimento estão descritos abaixo.

Quadro 53: Principais parâmetros das redes de MLP no experimento 12.

Tipos de Redes	MLP – Multilayer Perceptron
Tamanho da "janela" de entrada	5 elementos
Parada ("goal")	0,001
Taxa de aprendizado	0,01
Épocas	2.000
Funções de ativação	Tansig, tansig e purelin
Função de treinamento	RPROP - Resilient Backpropagation
Normalização dos dados	Reg = $2 ./ (1 + \exp(-0.004*n)) - 1$ MM05, MM10, MM15, MM20
Quantidade total de registros	1040
Dados de treinamento	728
Dados de teste	312

No item 14, do Anexo, têm-se os resultados alcançados para as agências 02, 06 e 13.

Verifica-se que mesmo com o uso de médias móveis o resultado não se mostrou promissor. Entretanto, cabe esclarecer que o amortecimento promovido nos dados fez com que a taxa de acerto no treinamento fosse superior a 85% mas, quando da retirada da média móvel dos dados o índice cai a patamares baixíssimos. Para melhor exemplificar, segue figura com o procedimento adotado.

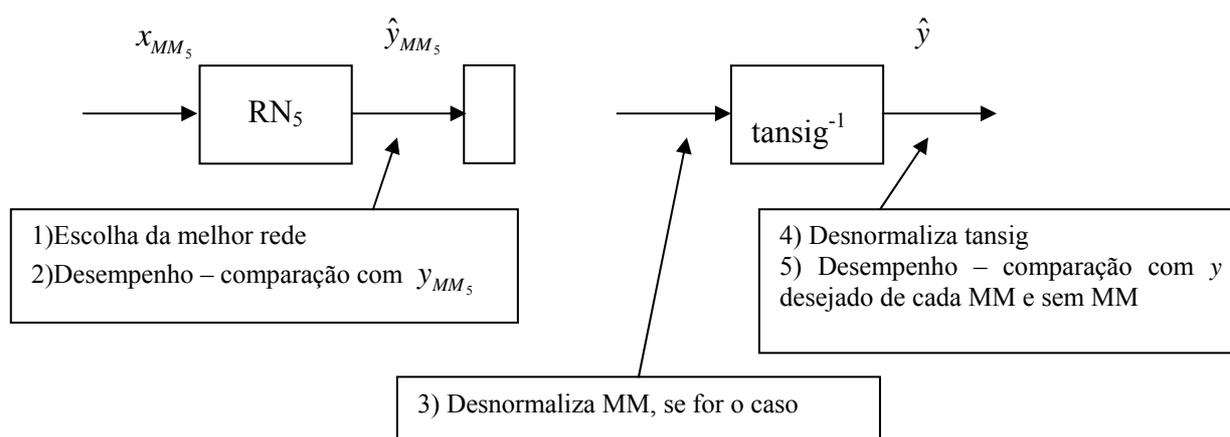


Figura 16: Estratégia de treinamento com o uso de médias móveis.

Na avaliação do desempenho (ponto 2 da figura), ou seja, a comparação dos dados gerados pela rede com os registros de treinamento, o índice de acerto ficou acima

de 85% demonstrando que a rede conseguiu mapear corretamente os dados. Entretanto, quando da medição da performance no ponto 5, logo após retirar a média móvel e desnormalizar os dados este mesmo índice cai para algo em torno de 10%.

7.14. MODELO NEBULOSO

Completando o trabalho, foram criados modelos nebulosos para verificar a performance e se comparar com os demais modelos (Arima, amortecimento exponencial e de redes neurais artificiais). Utilizou-se, neste caso, os mesmos parâmetros dos modelos anteriores, ou seja, considerou-se como corretas as previsões que tivessem valores no intervalo de $\pm 20\%$ do real.

Foram gerados, no total, 4 (quatro) modelos nebulosos a fim de se verificar qual o que melhor responderia aos dados. Para tanto, os dias úteis da semana foram divididos em 3 e 5 conjuntos e os dias do mês em 3 e 11 conjuntos. O mês do ano não foi *fuzzificado* e cada mês correspondeu a um valor entre 1 e 12, iniciando-se com janeiro. O dia da semana e o dia do mês foram *fuzzificados* de duas formas cada um e combinados entre si. Assim, ficou-se com os seguintes modelos:

- Modelo 1: dias da semana divididos em 3 conjuntos; dias do mês divididos em 3 conjuntos e meses do ano correspondentes a 12 conjuntos. Neste modelo foram geradas 108 regras (3 x 3 x 12).
- Modelo 2: dias da semana divididos em 5 conjuntos; dias do mês divididos em 3 conjuntos e meses do ano correspondentes a 12 conjuntos. Neste modelo foram geradas 180 regras (5 x 3 x 12).
- Modelo 3: dias da semana divididos em 3 conjuntos; dias do mês divididos em 11 conjuntos e meses do ano correspondentes a 12 conjuntos. Neste modelo foram geradas 396 regras (3 x 11 x 12).
- Modelo 4: dias da semana divididos em 5 conjuntos; dias do mês divididos em 11 conjuntos e meses do ano correspondentes a 12 conjuntos. Neste modelo foram geradas 660 regras (5 x 11 x 12).

Por fim, aplicou-se em cada um dos modelos janelas de 1 a 10 posições.

A aplicação do modelo nebuloso nas agências 02, 06 e 13 teve os resultados mostrados no item 15, do Anexo.

Os valores apresentados não se mostraram satisfatórios apresentando índices de acerto extremamente baixos. Acredita-se que este fato é decorrente da oscilação verificada na base, dificultando a obtenção das características da função.

7.15. CONCLUSÃO

Nenhum dos modelos tratados – ARIMA, RNA e *fuzzy* – respondeu bem aos experimentos. Cabe ressaltar que a idéia de se adotar o intervalo de $\pm 20\%$ como meta pode não ser a melhor opção pois para valores muito pequenos o intervalo para considerar a previsão como correta é muito restrito. Além disso como as agências possuem a maioria dos dados próximos de 0 (zero) a aplicação do percentual acima citado faz com que previsões que poderiam ser consideradas boas para a aplicação prática na definição do valor necessário para o funcionamento da agência tenham sido descartadas em função de haver sido extrapolado os 20% definidos. Por exemplo: uma previsão de R\$ 13 mil seria considerada errada se o valor real fosse R\$ 10 mil pois o intervalo definido para se considerar o valor previsto como correto seria entre R\$ 8 mil e R\$ 12 mil. Entretanto, para efeito prático esse “erro” é considerado desprezível pelos analistas do Banco, não trazendo maiores conseqüências para o funcionamento da agência bancária.

Por outro lado, nos pontos cujo valor é grande (por exemplo, acima de R\$ 200 mil) o fato de se prever “corretamente” pode não ser útil para a agência pois o dado previsto pode estar muito distante do real e ainda assim ser considerado correto. Temos como exemplo o dia 06/01/1998, da agência 06, cujo valor real é de R\$ 2.905.640,00. Neste caso seria considerada correta uma previsão que se situasse no intervalo [2.324.510,00 3.486.770,00]. Obviamente que apesar de um valor previsto de 3.480.000,00 ser considerado correto traria grandes riscos para o Banco pois teria-se um erro de R\$ 574.360,00.

8. ESTUDO DA ALEATORIEDADE DA SÉRIE

Este capítulo trata dos conceitos e técnicas empregadas no estudo da aleatoriedade. Na seção 1 discorre-se sobre amostragem aleatória para, em seguida, nos itens 2 a 5 apresentar os métodos que foram aplicados à série. Em seguida, na seção 6, têm-se os resultados obtidos e, por último, as conclusões.

8.1. AMOSTRAGEM ALEATÓRIA

No caso de populações discretas uma amostra aleatória é aquela em que cada item da população tem a mesma chance de ser incluído na amostra. Já nos casos de populações contínuas uma amostra aleatória é aquela em que a probabilidade de incluir na amostra qualquer intervalo de valores é igual à percentagem da população que está naquele intervalo (STEVENSON, 1986, p. 161).

Os números aleatórios podem ser classificados em 3 grupos:

- Aleatórios verdadeiros: são aqueles que exibem um comportamento verdadeiramente aleatório, ou seja, não apresentam qualquer correlação entre eles. São selecionados por meio não determinístico e não envolvem o cálculo de funções ou o uso de algoritmos. Geralmente são obtidos por intermédio de eventos físicos ou naturais;
- Pseudo-aleatórios: são obtidos por meio de algoritmos, e procuram apresentar um ciclo de repetição tão alto quanto possível, de modo a simular uma distribuição verdadeiramente randômica. Possuem um aspecto aleatório mas, contudo, apresentam um padrão específico e repetitivo;
- Quase-aleatórios: conhecidos como seqüências de baixa discrepância, são obtidos por procedimentos matemáticos. Discrepância é um conceito na teoria dos números que mede o desvio da uniformidade de um conjunto de pontos em uma dimensão d .

Segundo Wasserman (NETER; WASSERMAN; WHITMORE, 1992, p. 245), em uma população infinita, um conjunto de variáveis X_1, X_2, \dots, X_n gerado por

um processo constitui uma amostra randômica simples se as variáveis são independentes e têm a mesma probabilidade de distribuição. Por variáveis independentes entende-se que as seqüências devem ser não correlacionadas, ou seja, qualquer subconjunto de números aleatórios não deve estar correlacionado com nenhum outro subconjunto da mesma população. Relativamente a se possuir a mesma probabilidade de distribuição, tem-se que a seqüência de números deve ser uniforme e não polarizada, ou seja, regiões iguais do espaço devem ser preenchidas por quantidades idênticas de números aleatórios.

A fim de avaliar se um conjunto de números pode ser considerado como aleatório, várias técnicas foram definidas. Wasserman (NETER; WASSERMAN; WHITMORE, 1992, p. 459) sugere que primeiramente se plotem gráficos do conjunto assim como se criem histogramas para verificar a distribuição dos dados. Entretanto, apesar dos gráficos auxiliarem na identificação de distorções da série aleatória eles não conseguem mensurar o quão próximo ou afastado este se encontra de um conjunto verdadeiramente aleatório e, para suprir esta lacuna, sugere o caso de testes estatísticos.

Dentre os diversos métodos existentes na literatura, optou-se pelo uso de 5 técnicas para avaliação do grau de aleatoriedade das séries sob estudo, quais sejam:

- Teste da autocorrelação;
- Teste do qui-quadrado;
- Teste de Kolmogorov-Smirnov; e
- Teste de corrida.

A título de estudo comparativo, inseriu-se na análise mais 2 séries onde a primeira delas corresponde ao movimento médio da cotação da ação PN (Preferencial Nominativa) da Petrobrás, código Bovespa Petr4, na bolsa de valores de São Paulo (SP), ocorrido no mesmo período da série das agências, ou seja, de 02/01/1998 à 28/03/2003, perfazendo um total de 1.044 registros e a segunda correspondendo a 1.059 pontos gerados através da função *randn*, do aplicativo MatLab.

8.2. MÉTODO DA AUTOCORRELAÇÃO

A função de autocorrelação corresponde à correlação interna entre elementos de uma série de observações ordenadas no tempo ou no espaço. A força de uma relação entre duas variáveis em uma população é geralmente medida pelo

coeficiente de correlação, cujos valores abrangem desde -1 para a correlação negativa perfeita até $+1$ para a correlação positiva perfeita e é dado por:

$$r = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_x} \right) \left(\frac{y_i - \bar{y}}{\sigma_y} \right) \quad \text{Eq. (67)}$$

onde n = número de elementos da amostra;

x_i, y_i = elementos no ponto i ;

\bar{x}, \bar{y} = média aritmética dos elementos x e y , respectivamente; e

σ_x, σ_y = desvio padrão dos elementos x e y , respectivamente.

Entretanto, é possível correlacionar os valores de uma variável x , em certas épocas, com os correspondentes x em ocasiões anteriores. Isso é freqüentemente denominado autocorrelação e a equação descritiva desta função encontra-se a seguir.

$$r = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x})(x_{i+j} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{Eq. (68)}$$

onde n = número de elementos da amostra;

x_i = elementos no ponto i ;

\bar{x} = média aritmética dos elementos x ;

j = intervalo de tempo.

O cálculo da autocorrelação indica, nos casos em que o coeficiente é alto, que os valores da série tendem a formar seqüências equivalentes de tempos em tempos, produzindo uma ciclicidade, ou seja, a série é caracterizada por seqüências repetitivas em intervalos pré-definidos. Assim, uma amostra onde o coeficiente de autocorrelação é nulo indica que os dados são independentes e identicamente distribuídos.

Um dos primeiros testes empíricos para a avaliação da eficiência do mercado financeiro foi o de autocorrelação serial e este teste serviu para avaliar o grau de dependência entre as taxas de rentabilidade de um dia com as taxas de rentabilidade de dias anteriores.

Uma autocorrelação serial igual a zero significa que as mudanças de preços em períodos de tempos consecutivos não são correlacionáveis entre si e, dessa forma, os investidores não poderiam obter retornos extraordinários a partir de informações passadas (SILVA, 2004).

8.3. MÉTODO DO QUI-QUADRADO

Os resultados obtidos por meio de amostras nem sempre concordam exatamente com os teóricos esperados, de acordo com as regras de probabilidade. Por exemplo: suponha que, em uma determinada amostra, observou-se que um conjunto de eventos possíveis E_1, E_2, \dots, E_n ocorreram com as freqüências x_1, x_2, \dots, x_n denominadas freqüências observadas e, de acordo com as regras de probabilidade, esperava-se que elas ocorressem com as freqüências e_1, e_2, \dots, e_n , denominadas freqüências esperadas ou teóricas. Dessa forma deseja-se, na maioria das vezes, saber se as freqüências observadas diferem, de modo significativo, das esperadas.

Uma medida de discrepância existente entre as freqüências observadas e esperadas é proporcionada pela estatística X^2 (qui-quadrado) e foi proposta, em 1900, por Karl Pearson, sendo considerado um dos melhores testes estatísticos. Este método compara as freqüências observadas x_1, x_2, \dots, x_n , de k valores (ou em k intervalos) de uma variável, com as correspondentes freqüências e_1, e_2, \dots, e_n de uma distribuição teórica esperada e é expresso, matematicamente, por:

$$X^2 = \sum_{i=1}^k \frac{(x_i - e_i)^2}{e_i} \quad \text{Eq. (69)}$$

onde k = número de intervalos em que foi subdivida a amostra;

x_i = freqüências observadas;

e_i = freqüências esperadas.

Quando $X^2 = 0$, as freqüências teóricas e observadas concordam exatamente, enquanto que, quando $X^2 > 0$ isso não se dá. Quanto maior for o valor de X^2 , maior será a discrepância entre as freqüências observadas e esperadas, ou seja, o teste do qui-quadrado pode ser usado para determinar quanto aproximadamente às distribuições teóricas se ajustam às distribuições empíricas, isto é, as obtidas por meio de dados amostrais (SPIEGEL, 1961, p. 332-333).

Por exemplo: Deseja-se verificar se há associação entre 3 cursos de uma universidade e se os alunos possuem veículo de transporte próprio. A tabela de distribuição de freqüências está a seguir (MORCILLO, 2004).

Quadro 54: Distribuição de frequência de propriedade de automóvel por curso.

	Medicina	Direito	Engenharia	Total
Possui automóvel	10	20	30	60
Não possui automóvel	15	15	30	60
Total	25	35	60	120

Calcula-se, primeiramente, a frequência esperada de cada célula através da seguinte fórmula

$$E = \frac{T_c \cdot T_l}{N} \quad \text{Eq. (70)}$$

onde T_c = Total da coluna;

T_l = Total da linha;

N = Total de observações.

As frequências esperadas são anotadas nas correspondentes células

Quadro 55: Frequências observadas e esperadas.

		Medicina	Direito	Engenharia	Total
Possui automóvel	Observado	10	20	30	60
	Esperado	12,5	17,5	30	
Não possui automóvel	Observado	15	15	30	60
	Esperado	12,5	17,5	30	
Total		25	35	60	120

A seguir aplica-se a fórmula do qui-quadrado, para os dados constantes do quadro acima.

$$X^2 = \frac{(10-12,5)^2}{12,5} + \frac{(20-17,5)^2}{17,5} + \frac{(30-30)^2}{30} + \frac{(15-12,5)^2}{12,5} + \frac{(15-17,5)^2}{17,5} + \frac{(30-30)^2}{30} = 1,7$$

Determinam-se os graus de liberdade (gl) da tabela.

Os graus de liberdade da tabela são calculados multiplicando-se o número de linhas da tabela menos um pelo número de colunas da tabela menos um.

$$gl = (\text{número de linha} - 1) \times (\text{número de colunas} - 1)$$

No exemplo, tem-se:

$$gl = (3 - 1) \times (2 - 1) = 2$$

Por último, compara-se o valor do X^2 obtido como o X^2 crítico, que se encontra em tabelas, considerando os graus de liberdade e o nível de significância.

Neste caso o $X^2 = 1,7$ é menor que o valor constante da tabela que é de 5,99. Assim sendo, a hipótese H_0 não pode ser rejeitada, concluindo-se que no grupo estudado não há associação entre as variáveis.

8.4. MÉTODO DE KOLMOGOROV-SMIRNOV

Uma das maneiras de verificar se uma distribuição teórica se ajusta ao fenômeno investigado ou não é através de testes de aderência. Através de funções empíricas e certas tolerâncias definidas pelo usuário, estes testes comparam a distribuição teórica com a aproximada para cada intervalo do diagrama de freqüências relativas e, no final, dizem, de acordo com nível de confiança pré-estabelecido, se a distribuição teórica pode representar o fenômeno. Dentre estes testes estão o teste do qui-quadrado e o de Kolmogorov-Smirnov (K-S) (LIMA; SAGRILO, 2003, p. 27).

O princípio do teste de K-S baseia-se na comparação da curva da freqüência cumulativa dos dados, com a função de distribuição teórica em hipótese. Quando as duas curvas se sobrepõem à estatística de teste é calculada através da máxima diferença entre ambas. A magnitude da diferença estabelece-se probabilisticamente, segundo a lei de probabilidade dessa estatística, que se encontra tabelada. Se os dados experimentais se afastam significativamente do que é esperado segundo a distribuição em hipótese, então as curvas obtidas devem encontrar-se igualmente afastadas, e por um raciocínio análogo, se o ajustamento ao modelo hipotético é admissível, as curvas têm um delineamento próximo (NEVES, 2003, p. 2.5).

O procedimento básico do teste de Kolmogorov-Smirnov (K-S) é a comparação entre a freqüência acumulada experimental e a função de distribuição teórica assumida. Se a discrepância é grande em relação ao que seria normalmente esperada para uma dada amostra, o modelo teórico é rejeitado.

Para uma amostra de tamanho n , os dados devem ser organizados em ordem crescente, e a partir deles monta-se a função de freqüência acumulada da seguinte forma:

$$S_n(x) = \begin{cases} 0 & x < x_1 \\ \frac{k}{n} & x_k \leq x < x_{k+1} \\ 1 & x \geq x_n \end{cases} \quad \text{Eq. (71)}$$

onde x_1, x_2, \dots, x_n são os valores dos dados amostrais ordenados, e n é o tamanho da amostra. A figura abaixo mostra os valores plotados de $S_n(x)$ e também a função de distribuição teórica proposta $F(x)$.

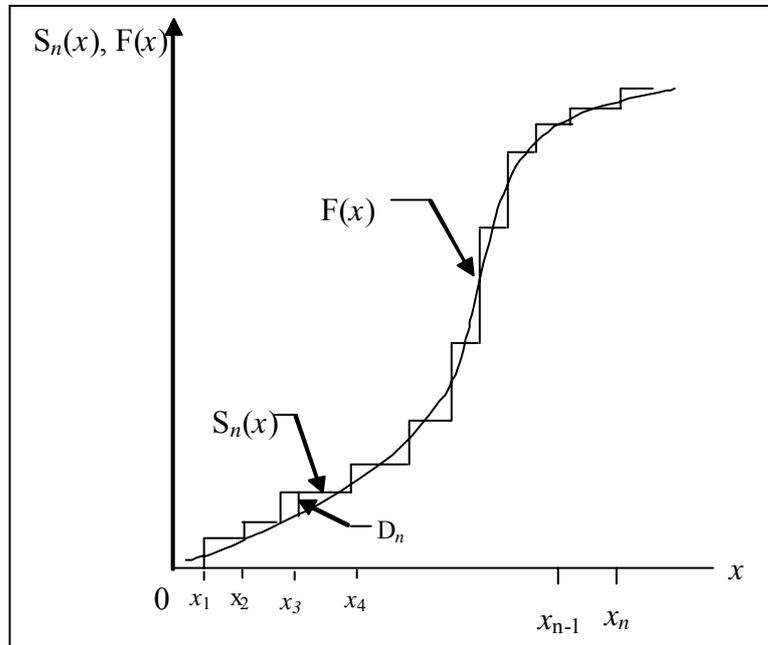


Gráfico 49: Representação da frequência experimental acumulada $S(x)$ e função teórica proposta $F(x)$

No teste K-S, a diferença máxima entre $S_n(x)$ e $F(x)$ para todos os valores de X é a medida de discrepância procurada. Este valor máximo é:

$$D_n = \max | F(x) - S_n(x) | \quad \text{Eq. (72)}$$

Para um nível de significância α , o teste K-S compara a diferença máxima observada D_n com o valor crítico D_n^α , que é definido como:

$$P(D_n \leq D_n^\alpha) = 1 - \alpha \quad \text{Eq. (73)}$$

Os valores críticos de D_n^α para vários níveis de significância são tabelados para diferentes valores de n e podem ser encontrados em um bom livro de estatística (NETER; WASSERMAN; WHITMORE, 1992, p. 937)(AFIF; AZEU, 1979, p. 414). Se o valor de D_n observado for menor que o valor crítico, então a distribuição proposta é aceitável naquele nível de significância específico; caso contrário, a distribuição assumida deve ser rejeitada (EBECKEN, 2001, p. 39-40).

Por exemplo: Sortearam-se ao acaso dois conjuntos com 45 elementos com a expectativa que tivessem média 100 e desvio padrão 15. Como este sorteio foi realizado utilizando um software, deseja-se verificar se há diferença entre as duas amostras.

Quadro 56: Teste de Kolmogorov-Smirnov – Valores sorteados.

Grupo 1			Grupo 2		
92	110	103	113	95	83
106	100	81	77	105	81
125	113	92	98	121	73
101	123	106	104	79	89
111	96	97	106	121	92
80	75	104	121	118	102
84	91	108	83	65	111
85	112	64	108	76	116
97	93	96	103	106	93
104	77	95	101	107	83
81	93	122	123	109	111
129	104	102	98	89	114
121	85	104	93	97	78
89	116	92	78	108	110
81	97	106	105	131	98

Para os dados acima têm-se as seguintes hipóteses:

H_0 : o grupo 1 e o grupo 2 têm a mesma distribuição

H_1 : a distribuição entre os grupos 1 e 2 é diferente.

Os procedimentos a serem adotados para cálculo do teste são:

1. Estabelecer os intervalos de classe.
2. Determinar as frequências dentro de cada intervalo.
3. Determinar as frequências acumuladas dentro de cada intervalo.
4. Calcular as proporções das frequências acumuladas dentro de cada intervalo de classe.
5. Calcular as diferenças entre a proporção do grupo 1 e do grupo 2 dentro de cada intervalo de classe.
6. Selecionar D.
7. Calcular $D_{\text{crítico}}$ e tomar a decisão.

O quadro adiante apresenta os dados compilados após os cálculos dos passos 1 a 5 acima.

Quadro 57:Resumo dos procedimentos 1 a 5 do teste de Kolmogorov-Smirnov

	Grupo 1			Grupo 2			p1-p2
	f	freqüência acumulada	p1	f	freqüência acumulada	p2	
60 - 65	1	1	0,022	1	1	0,022	0,000
65 - 70	0	1	0,022	0	1	0,022	0,000
70 - 75	1	2	0,044	1	2	0,044	0,000
75 - 80	2	4	0,089	5	7	0,156	-0,067
80 - 85	5	9	0,200	4	11	0,244	-0,044
85 - 90	2	11	0,244	2	13	0,289	-0,045
90 - 95	7	18	0,400	4	17	0,378	0,022
95 - 100	6	24	0,533	4	21	0,467	0,066
100 - 105	7	31	0,689	6	27	0,600	0,089
105 - 110	5	36	0,800	7	34	0,756	0,044
110 - 115	3	39	0,867	4	38	0,844	0,023
115 - 120	1	40	0,889	2	40	0,889	0,000
120 - 125	4	44	0,978	4	44	0,978	0,000
125 - 130	1	45	1,000	0	44	0,978	0,022
130 - 135	0	45	1,000	1	45	1,000	0,000
Total	45			45			

A diferença entre as proporções de freqüências p1 e p2 com maior valor absoluto ocorreu na classe [100 -| 105], portanto, o $D_{\text{máx}} = |0,089| = + 0,089$.

Considerando um índice de significância $\alpha = 0,05$, tem-se:

$$D_{\text{crítico};0,05} = 1,36 \cdot \sqrt{\frac{n1+n2}{n1 \cdot n2}} = 1,36 \cdot \sqrt{\frac{45+45}{45 \cdot 45}} = 1,36 \cdot 0,210818 = 0,2867$$

Como o valor do $D_{\text{máx}}$ calculado é menor que o $D_{\text{crítico}}$ não se pode rejeitar H_0 , significando que tanto o grupo 1 como o grupo 2 possuem a mesma distribuição.

8.5. MÉTODO DE CORRIDAS

A corrida se refere a uma sucessão de eventos similares separados por eventos diferentes, ou seja, uma seqüência ininterrupta de observações que apresentam uma característica semelhante. As corridas focalizam a atenção na seqüência temporal (ordenada no tempo), como o registro diário da temperatura de uma localidade, ou espacial (ordenada por posição), a exemplo da chegada de homens e mulheres em uma fila. Independentemente da natureza efetiva dos dados, as observações podem ser transformadas numa série de símbolos em que se possa contar o número de repetições (STEVENSON, 1986, p. 324-325).

8.5.1. Corrida

Corrida é uma sucessão não interrompida de dígitos da seqüência original. Uma corrida de comprimento k consiste em exatamente k dígitos, limitada antes e depois por outro segmento composto por valores opostos. Os testes de corrida verificam se a ocorrência de seqüências - crescentes e decrescentes, acima e abaixo etc - apresentam número e comprimento aleatórios

Por exemplo: Sejam as seqüências $|1\ 2\ 9|$, $|8\ 5\ 3\ 0|$ e $|7\ 8\ 9|$ a serem avaliadas. Neste caso têm-se duas corridas crescentes ($|1\ 2\ 9|$ e $|7\ 8\ 9|$) e uma corrida decrescente ($|8\ 5\ 3\ 0|$).

Complementando o conceito, dada a seqüência 0,123 0,357 0,789 0,425 0,224 0,416 0,950 as corridas são contabilizadas da seguinte forma: existe uma corrida crescente de comprimento 3, uma corrida decrescente de comprimento 2 e uma corrida crescente de (pelo menos) 2, dependendo dos próximos valores.

Quadro 58: Seqüência hipotética

0,123	0,357	0,789	0,425	0,224	0,416	0,950
Corrida crescente de comprimento 3			Corrida decrescente de comprimento 2		Corrida crescente de comprimento 2	

8.5.2. Análise das Repetições

Para podermos dizer que uma série de observações apresenta as características de um processo aleatório, é preciso primeiro ter uma idéia daquilo que um processo aleatório pode produzir. Isto servirá então de padrão de julgamento de uma situação efetiva. A distribuição amostral proporciona uma definição de aleatoriedade indicando o número esperado (médio) de repetições e revelando qual o grau de variabilidade razoável (para uma série aleatória) entre os números de repetições esperado e observado.

Quando a série em estudo envolve um grande número de observações (> 30) a distribuição amostral do número de repetições é bem aproximada pela distribuição normal.

A hipótese nula num teste de repetições é que a diferença entre o número esperado de repetições e o número efetivamente observado é o resultado de variação aleatória na amostragem. A hipótese alternativa é que a diferença entre o número

observado e o número esperado de repetições é demasiado grande para ser devida só ao acaso. Testa-se a diferença comparando-a com a variabilidade esperada das repetições, supondo que haja somente aleatoriedade nos dados. Isto é, divide-se a diferença entre os números observados e esperados de repetições pelo desvio padrão da distribuição adequada. Assim:

$$z = \frac{\text{número observado de repetições} - \text{número esperado de repetições}}{\text{desvio padrão das repetições}} \quad \text{Eq. (74)}$$

ou

$$z = \frac{r - R}{\sigma_R} \quad \text{Eq. (75)}$$

Pequenos valores de z (digamos, menor do que 1,96 para $\alpha = 0,05$, onde α corresponde ao índice de significância) sugerem que a seqüência é aleatória, enquanto grandes valores sugerem o contrário (STEVENSON, 1986, p. 327).

Uma repetição é uma seqüência ininterrupta de observações que possuem uma característica comum e focalizam a atenção na ordem de ocorrência das observações. As repetições podem constituir uma seqüência temporal (ordenada no tempo) ou uma seqüência espacial (ordenada por posição). Em ambos os casos usa-se o teste de repetições para detectar padrões não aleatórios em dados quando se conhece a ordem das observações. Por exemplo, a existência de uma seqüência do tipo ABCD ABCD ABCD ABCD ABCD ABCD sugere um ciclo, já uma seqüência do tipo AAAAAAAAAA BBBBBB sugere uma tendência de A para B nos dados.

Os testes de repetições se propõem a identificar a existência de algum padrão nos dados a fim de poder classificá-los como aleatórios ou não. Entretanto, pode-se ocorrer de na aplicação dos testes um deles, por exemplo o teste acima e abaixo da mediana, indicar uma não aleatoriedade de variabilidade, enquanto o outro indicar o contrário. Neste caso, a série deve ser considerada como não aleatória.

8.5.3. *Teste de repetições de dois tipos de observações*

Quando os dados em estudo são dicotomizados, ou seja, pertencem a duas categorias (verdadeiro-falso, positivo-negativo, etc), a definição de aleatoriedade do

conjunto deve ser feita após a contagem do número de repetições dos dados e comparação com o número esperado de repetições.

A distribuição amostral para o número de repetições com dois tipos de observação é aproximadamente normal para grandes amostras (> 20) e o número esperado de repetições, para esta situação, é dado por:

$$R = \frac{2n_1n_2}{N} + 1 \quad \text{Eq. (76)}$$

onde R = número esperado de repetições;

n_1 = número de observações de um tipo;

n_2 = número de observações de outro tipo; e

N = número total de observações.

O desvio padrão da distribuição amostral do número de repetições é dado por:

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}} \quad \text{Eq. (77)}$$

Que, após alguns procedimentos, pode ser reduzido à equação abaixo:

$$\sigma_R = \sqrt{\frac{(R-1)(R-2)}{N-1}} \quad \text{Eq. (78)}$$

Por exemplo: Têm-se quarenta pessoas em uma fila de banco aguardando o atendimento. Pode-se dizer que a seqüência é aleatória? (STEVENSON, 1986, p. 327-328).

Seja M a designação das Mulheres e H a designação dos Homens. A seqüência é:

MMMMM HH MM HH MMMM VVV MMMMMM HHHHH M HHH MMMMMM H

As repetições estão separadas por espaço. Chamando n_1 o número de M 's e n_2 o número de H 's, vemos que $n_1 = 24$ e $n_2 = 16$. Há 12 repetições ($r = 12$). Logo, o número esperado de repetições é:

$$R = \frac{2n_1n_2}{N} + 1 = \frac{2(24)(16)}{40} + 1 = 19,2 + 1 = 20,2$$

O desvio padrão é dado por:

$$\sigma_R = \sqrt{\frac{(R-1)(R-2)}{N-1}} = \sqrt{\frac{(19,2)(18,2)}{39}} = 2,99$$

Usando o desvio padrão como medida da diferença relativa entre o número efetivo e o número esperado de repetições, vemos que:

$$z = \frac{r - R}{\sigma_R} = \frac{12 - 20,2}{2,99} = \frac{-8,2}{2,99} \approx -2,7$$

Como se vê na figura abaixo, a diferença z está acima do que poderia ser considerado dentro de uma distribuição uniforme $\pm 1,96$. Conclui-se, então, que a seqüência não é aleatória.

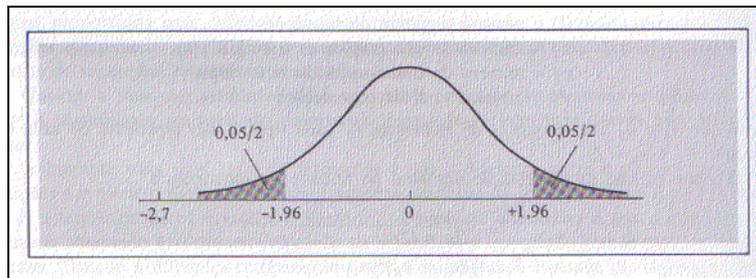


Figura 17: Representação da normal com os limites críticos para o nível de significância 0,01.

8.5.4. Teste de repetições acima e abaixo da mediana

Quando os dados são numéricos, a detecção da aleatoriedade de uma série pode ser feita em relação à mediana amostral. As medições são caracterizadas como maiores (M) ou menores (m) em relação à mediana e ignoram-se os valores coincidentes. Contam-se, então, os números de repetições M e m.

A conversão de cada valor para M ou m tornará o problema com dois tipos de observações apenas (dicotomizado) podendo-se, então, utilizar as fórmulas anteriores. Entretanto, com o número de M's igualará o de m's (por definição, metade das observações está acima da mediana e metade abaixo), as fórmulas anteriores ficam simplificadas (STEVENSON, 1986, p. 328-329):

$$R = \frac{N}{2} + 1 \quad \text{Eq. (79)}$$

$$\sigma_R \approx \sqrt{\frac{(N-1)}{4}} \quad \text{Eq. (80)}$$

onde N corresponde ao número total de observações.

Por exemplo: A divisão de saúde de uma cidade registra a quantidade de dióxido de enxofre na atmosfera durante 44 dias sucessivos. Que se pode concluir quanto à aleatoriedade da seqüência? Pelos registros históricos a mediana equivale a 0,050. A seguir tem-se o quadro com as medições realizadas na localidade estudada.

Quadro 59: Quantidade de dióxido de enxofre na atmosfera durante 44 dias sucessivos.

Dia	SO ₂		Dia	SO ₂	
1	0,057	A	22	0,051	A
2	0,040	B	23	0,063	A
3	0,059	A	24	0,060	A
4	0,063	A	25	0,049	B
5	0,061	A	26	0,040	B
6	0,040	B	27	0,044	B
7	0,009	B	28	0,058	A
8	0,003	B	29	0,032	B
9	0,031	B	30	0,018	B
10	0,067	A	31	0,017	B
11	0,071	A	32	0,017	B
12	0,083	A	33	0,030	B
13	0,081	A	34	0,053	A
14	0,093	A	35	0,054	A
15	0,065	A	36	0,085	A
16	0,023	B	37	0,081	A
17	0,029	B	38	0,041	B
18	0,018	B	39	0,037	B
19	0,001	B	40	0,063	A
20	0,010	B	41	0,073	A
21	0,055	A	42	0,055	A
22	0,056	A	43	0,048	B

Os dados apresentados mostram a existência de 14 repetições. Se a seqüência é aleatória, deve-se esperar

$$R = \frac{N}{2} + 1 = \frac{44}{2} + 1 = 23 \text{ repetições}$$

E o desvio padrão da distribuição amostral correspondente será

$$\sigma_r = \sqrt{\frac{N-1}{4}} = \sqrt{\frac{43}{4}} = \sqrt{10,75} \approx 3,28$$

Deseja-se saber, então, se um processo realmente aleatório pode ocasionar apenas 14 repetições, quando o número esperado é de 23, conforme cálculo de R acima. Usando o desvio padrão como medida da diferença entre os dois, tem-se

$$z = \frac{r - R}{\sigma_r} = \frac{14 - 23}{3,28} = \frac{-9,0}{3,28} \approx -2,74$$

Ao nível de 0,01 H_0 seria rejeitada e conclui-se que a seqüência não é aleatória. A figura abaixo ilustra esta situação.

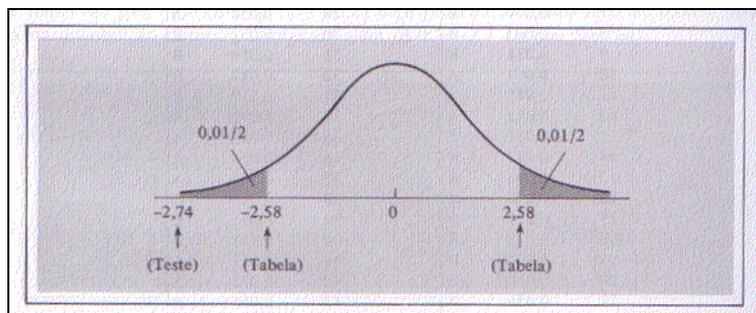


Figura 18: Representação da normal com os limites críticos para o nível de significância 0,01.

8.5.5. Teste de repetições acima e abaixo em relação ao anterior

Por fim, tem-se o teste de repetições para cima e para baixo como medida de aleatoriedade. Mede-se cada valor da seqüência em relação ao valor anterior e cada valor numérico é substituído por U (para cima) e D (para baixo), em relação ao valor anterior. Cabe lembrar que ao primeiro valor não se atribui símbolo pois não possui antecedente.

Para grandes amostras (≥ 40) a distribuição amostral do número de repetições é normal. Entretanto, as fórmulas da média e do desvio padrão da distribuição amostral não são as mesmas. A razão é que os testes anteriores envolveram elementos com probabilidade constante, como, por exemplo no caso de se ter um valor acima ou abaixo da mediana, cuja probabilidade de se obter um valor de um ou outro lado da mediana é 0,5 e não depende da observação anterior. Não é o caso das repetições para cima e para baixo. Se a seqüência é realmente aleatória, a probabilidade de um valor alto ser seguido por outro ainda mais alto é muito menor que 0,5, e a probabilidade de um valor elevado ser seguido por um menor é maior que 0,5. As

fórmulas, portanto, levam em conta esse aspecto de probabilidade “posicional” e o número esperado de repetições para cima e para baixo fica representado por (STEVENSON, 1986, p. 330-331):

$$R = \frac{2N - 1}{3} \quad \text{Eq. (81)}$$

onde N é o total de observações. O desvio padrão do número de repetições para cima e para baixo é:

$$\sigma_R = \sqrt{\frac{(16N - 29)}{90}} \quad \text{Eq. (82)}$$

Por exemplo: Seja a seqüência abaixo produzida por um gerador de números aleatórios e deseja-se conhecer a confiabilidade deste gerador, através de testes de aleatoriedade.

Quadro 60: Seqüência aleatória criada por um gerador de números randômicos.

76	88	1	35	34	49	17	89	19	41
	U	D	U	D	U	D	U	D	U
<hr/>									
14	99	13	23	79	40	15	19	1	66
D	U	D	U	U	D	D	U	D	U
<hr/>									
33	31	15	16	54	3	11	93	78	87
D	D	D	U	U	D	U	U	D	U
<hr/>									
50	23	46	14	27	12	38	12	20	15
D	D	U	D	U	D	U	D	U	D

Se a seqüência é aleatória, o número esperado de repetições para cima e para baixo será:

$$R = \frac{2N - 1}{3} = \frac{2 \cdot (40) - 1}{3} = \frac{79}{3} \approx 26$$

E o desvio padrão é dado por:

$$\sigma_R = \sqrt{\frac{16N - 29}{90}} = \sqrt{\frac{16 \cdot (40) - 29}{90}} = \sqrt{\frac{611}{90}} \approx \sqrt{6,79} \approx 2,61$$

O número observado de repetições é 32, de modo que a diferença entre o número esperado e o observado é 6. A diferença relativa é:

$$z = \frac{r - R}{\sigma_R} = \frac{32 - 26}{2,61} = \frac{6}{2,61} = +2,30$$

Ao nível de 0,05 H_0 seria rejeitado mas ao nível de 0,01 seria aceito, conforme pode ser visto na figura abaixo.

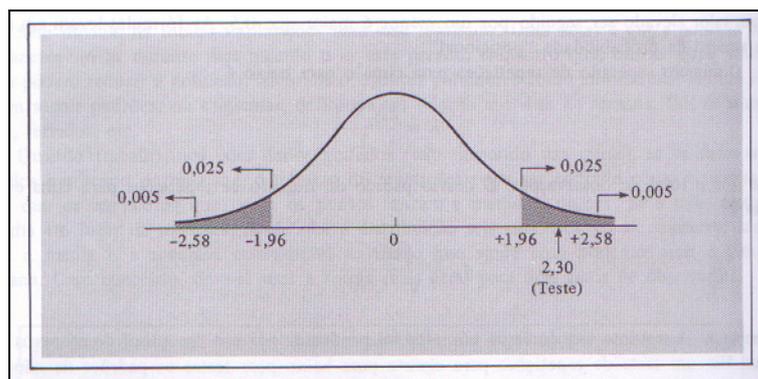


Figura 19: Representação da normal com os limites críticos para o nível de significância 0,01 e 0,05.

8.6. RESULTADOS

Os resultados obtidos através dos experimentos realizados com os diversos métodos estão compilados a seguir.

8.6.1. Teste da Autocorrelação

Neste teste foram calculadas as autocorrelações dos dados utilizando-se um deslocamento de 250 registros, ou seja, realizaram-se experimentos deslocando-se o primeiro dos vetores de 1 a 250 posições.

Abaixo têm-se as 10 maiores autocorrelações onde se vê que os dados relativos ao movimento médio das ações da Petrobrás possui índice de correlação relativamente alto, caracterizando uma causalidade ente os componentes da série. Já os dados randômicos gerados pelo aplicativo Matlab são os que têm os menores índices, como era de se esperar. Dentre as agências, a 13 apresentou os menores índices mas, ainda assim, altos para se considerar a série como sendo aleatória e baixos para caracterizar uma causalidade capaz de ser previsível e facilmente modelável.

Quadro 61: Os 10 maiores índices de correlação para as séries sob estudo.

	Maiores autocorrelações				
	Agência 02	Agência 06	Agência 13	Ações Petrobrás	Valores randômicos
1	0,428146	0,490699	0,252342	0,996848	0,087582
2	0,393657	0,329784	0,224469	0,992964	0,074217
3	0,357573	0,295445	0,221197	0,989302	0,067494
4	0,326892	0,287627	0,195052	0,985907	0,061429
5	0,31467	0,265159	0,18441	0,982601	0,057668
6	0,28983	0,264689	0,17432	0,97925	0,05712
7	0,287552	0,259759	0,16977	0,975782	0,056627
8	0,28307	0,258526	0,167163	0,972313	0,054655
9	0,280151	0,252264	0,164591	0,968697	0,051333
10	0,277688	0,221573	0,162907	0,965054	0,048303

Adiante têm-se os gráficos das autocorrelações de algumas das séries estudadas. Os valores foram plotados por deslocamentos, ou seja, o gráfico demonstra como a autocorrelação se comporta a cada deslocamento da série.

Analisando os gráficos pode-se perceber que as agências apresentam uma certa ciclicidade, notadamente nas agências 02. e 06. Isto nos leva a descartar a hipótese de aleatoriedade dos dados, pois a existência de ciclos significa previsibilidade no comportamento da série.

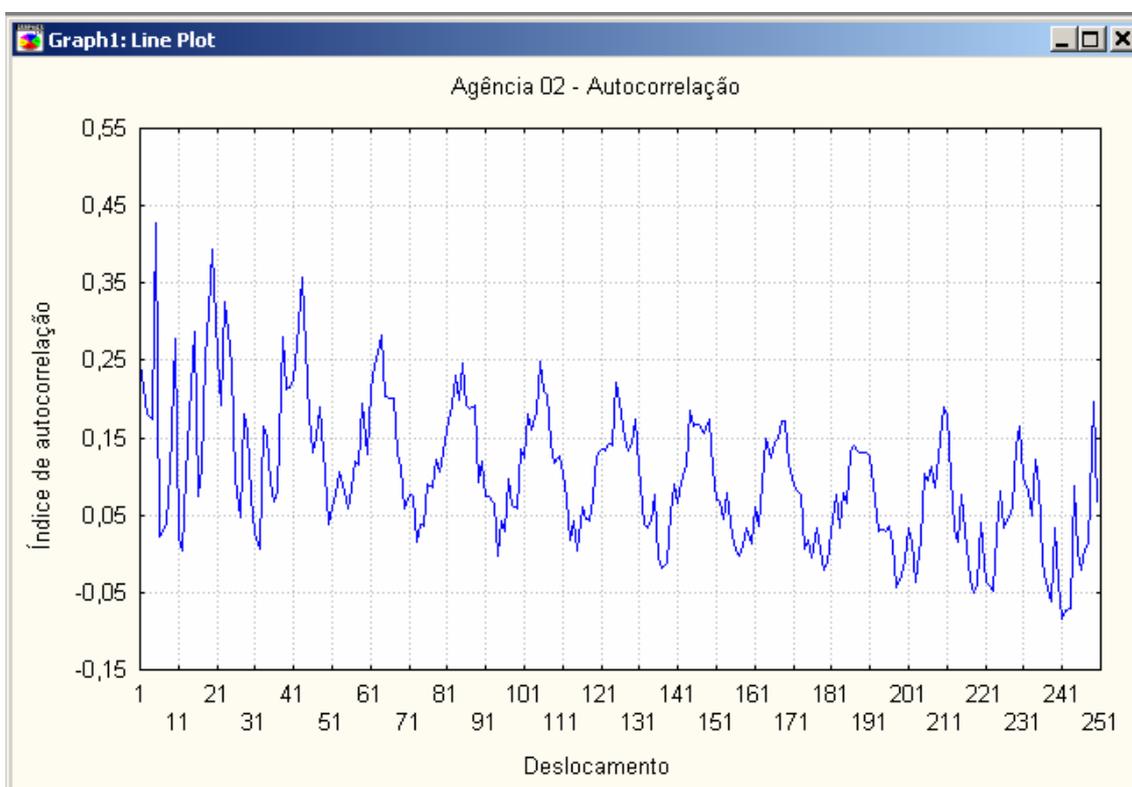


Gráfico 50: Agência 02 – Índices de autocorrelação por deslocamento.

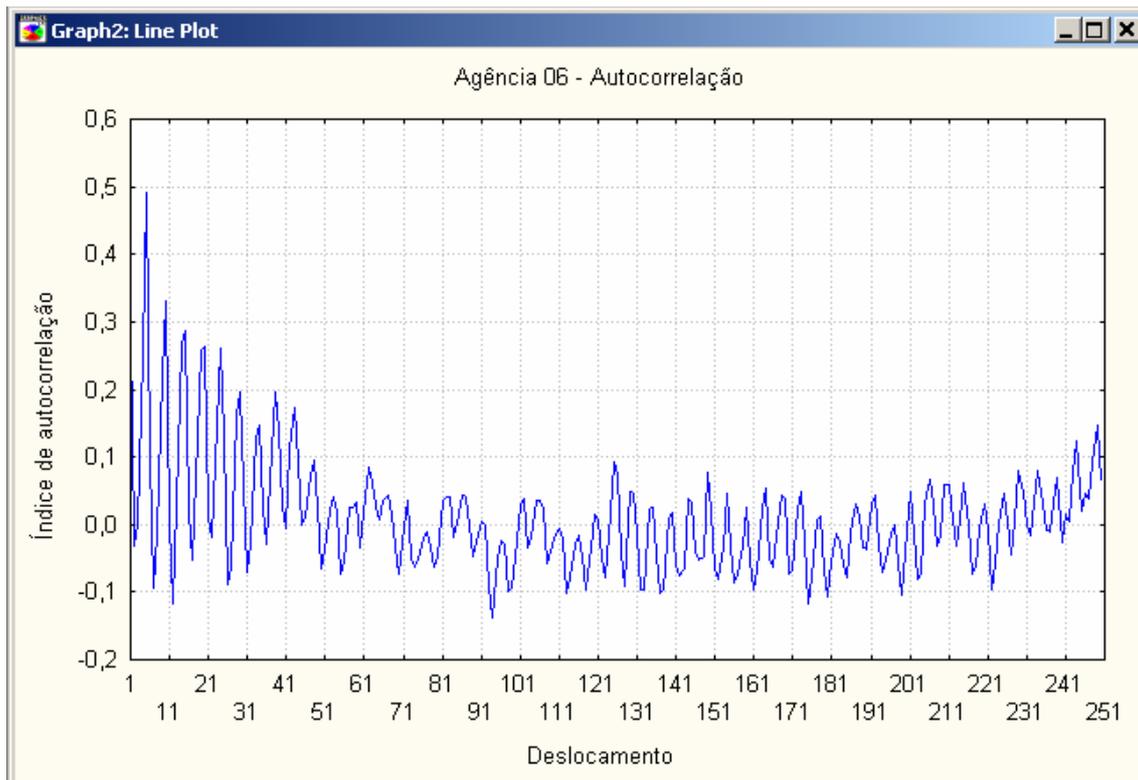


Gráfico 51: Agência 06 – Índices de autocorrelação por deslocamento.

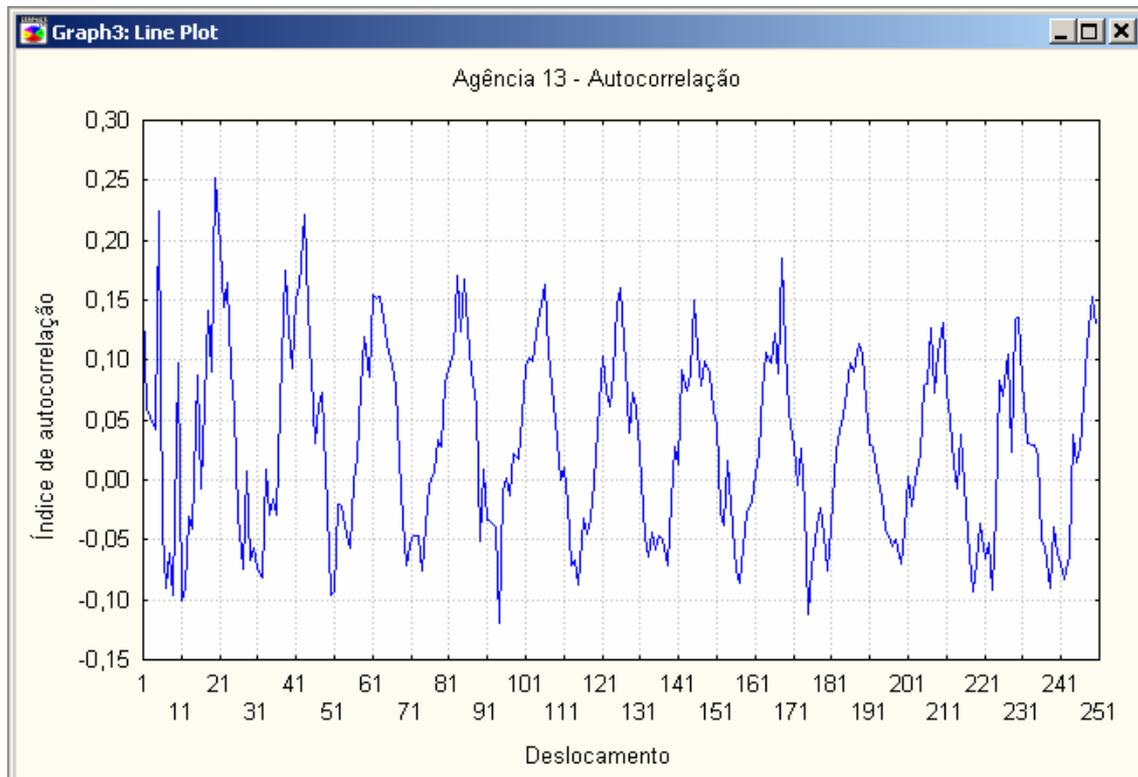


Gráfico 52: Agência 13 – Índices de autocorrelação por deslocamento.

Os gráficos relativos às séries aleatória e da Petrobrás não foram incluídos nesta etapa do estudo pois os índices calculados, vide quadro acima, já demonstram que se tratam de valores altamente correlacionados para o caso da Petrobrás (índice de autocorrelação próximo de 1) e descorrelacionados para a série aleatória (índice de autocorrelação próximo de 0).

8.6.2. Teste do qui-quadrado

O teste do qui-quadrado é utilizado para avaliar afirmações feitas sobre a distribuição de valores numa população. Conseqüentemente, é vantajoso dispor de um método que permita julgar se determinada população tem a distribuição exigida. Este teste avalia a associação existente entre variáveis qualitativas e compara as divergências entre as freqüências observadas e as esperadas.

Um dos primeiros procedimentos para se avaliar a aleatoriedade de um conjunto de dados consiste em provar que o conjunto é uniforme nos intervalos, ou seja, que a distribuição dos dados não segue a curva normal de probabilidades.

De uma maneira geral pode-se dizer que dois grupos se comportam de forma semelhante se as diferenças entre as freqüências observadas e as esperadas em cada categoria forem muito pequenas, próximas a zero.

Aplicando o método do qui-quadrado, foram divididos os dados em diversos intervalos de valores (de forma que contassem ao menos 5 observações em cada um, conforme recomendado para este teste), verificando-se o número de observações contidas. Em seguida calculou-se o número esperado de observações para cada intervalo, se a distribuição fosse autenticamente uma normal. A partir daí foram comparadas as diferenças entre os valores observados e os esperados por meio do teste.

O teste trabalha com duas hipóteses:

H₀: não há associação entre os grupos

H₁: há associação entre os grupos

Se, para um dado nível de significância, o valor do qui-quadrado for superior ao valor crítico da estatística do qui-quadrado (definido em tabela), rejeita-se a hipótese *H₀*.

No caso das séries estudadas (agências 02, 06 e 13, Petrobrás e série aleatória) foram utilizados vários intervalos de freqüência no cálculo (10, 20 e 50), ou

seja, distribuiu-se o conjunto de dados em 10, 20 e 50 grupos. Além disso, adotou-se o nível de significância igual a 0,05.

A exceção do intervalo de 50 posições para a série gerada através da função *randn* do Matlab, todos os demais cálculos foram rejeitados, conforme pode ser visto no quadro a seguir.

Quadro 62: Teste qui-quadrado para significância de 0,05 e várias frequências.

		Agência 02		Agência 06		Agência 13	
Freq.	Valor tabelado	Qui-quadrado	Situação registro	Qui-quadrado	Situação registro	Qui-quadrado	Situação registro
10	16,92	626,52	rejeitado	922,12	rejeitado	693,78	rejeitado
20	30,14	580,45	rejeitado	894,52	rejeitado	652,82	rejeitado
50	66,34	614,51	rejeitado	964,28	rejeitado	678,06	rejeitado

		Ações Petrobrás		Valores randômicos	
Freq.	Valor tabelado	Qui-quadrado	Situação registro	Qui-quadrado	Situação registro
10	16,92	444,14	rejeitado	54,29	rejeitado
20	30,14	553,24	rejeitado	37,07	rejeitado
50	66,34	566,34	rejeitado	50,11	aceito

Vê-se, portanto, que as séries não podem ser consideradas desassociadas, baseado no estudo proposto por Karl Pearson, ou seja, para um nível de confiança de 95% os números referentes às seqüências em estudo não são uniformemente distribuídos.

Cabe ressaltar que o fato da série randômica ter a sua hipótese aceita, para o último caso, não implica que seja aleatória pois outros testes têm que ser feitos a fim de se afirmar esta hipótese. Além disso, para os intervalos de 10 e 20 a hipótese foi rejeitada, significando que a seqüência gerada pelo Matlab não pode ser considerada randômica..

O fato dos valores das ações da Petrobrás serem inferiores aos das agências não significa, baseado neste teste, que esta série é mais aleatória que as demais. Simplesmente demonstra que a sua distribuição nos intervalos é mais uniforme que a das demais séries.

A literatura demonstra que para se considerar uma seqüência aleatória o conjunto de números deve ser aceito em todos os testes realizados. Caso um único experimento tenha sua hipótese rejeitada, deve-se descartar a suposição de serem randômicos.

Entretanto, esta é uma área (geração de números aleatórios - NA) em que vários trabalhos são realizados a fim de se obter geradores aleatórios eficazes e, os estudos mostram não ser uma tarefa trivial. Foi gerada uma seqüência de números randômicos utilizando-se uma função do Matlab (*randn*) que, a princípio, deveria apresentar um conjunto bom de dados e, no entanto, a série foi considerada ruim. Isto demonstra que, apesar das seqüências das agências 02, 06 e 13 não terem passado nos testes de aleatoriedade, não implica que sejam números com alguma previsibilidade, onde se encontre tendências ou ciclos bem definidos.

8.6.3. Teste de Kolmogorov-Smirnov

O princípio deste teste baseia-se na comparação da curva da freqüência cumulativa dos dados, com a função de distribuição teórica em hipótese. Quando as duas curvas se sobrepõem à estatística de teste é calculada através da máxima diferença entre ambas. A magnitude da diferença estabelece-se probabilisticamente, segundo a lei de probabilidade dessa estatística, que se encontra tabelada. Em outras palavras, o teste de Kolmogorov-Sminov mede a distância máxima entre os resultados de uma distribuição empírica a ser testada e os resultados associados à distribuição hipoteticamente verdadeira. Assim, o teste avalia se duas amostras têm distribuições semelhantes, ou seja, se foram extraídas de uma mesma população. (MORCILLO, 2004).

O princípio básico para realização do teste é comparar as freqüências acumuladas das duas amostras. Dessa forma, os seguintes passos devem ser adotados:

- Primeiramente ordenam-se as duas amostras.
- Constroem-se as distribuições de freqüências acumuladas nos intervalos de classe de cada amostra.
- Calculam-se as diferenças entre as freqüências acumuladas de cada amostra em cada um dos intervalos de classe. Este capítulo está estruturado sempre considerando as diferenças entre a primeira e a segunda amostra (A-B).
- Escolhe-se a maior diferença $[D_{máx}]$ que será comparada com $D_{crítico}$ (valor este tabelado).
- Se $D_{máx} \geq D_{crítico}$ rejeita-se a hipótese de igualdade das amostras.

Abaixo estão os resultados encontrados para o teste de Kolmogorov-Sminorv aplicado sobre os dados das séries em estudo.

O valor tabelado para a série das ações da Petrobrás é diferente dos demais em função dela possuir menos dados que as demais, ou seja, apenas 1044 registros. Isto ocorreu em decorrência de feriados locais onde, às vezes, a agência funciona e a Bolsa de Valores não. Cabe destacar que o período utilizado na série da Petrobrás é exatamente o mesmo daquele utilizado pelas agências.

Quadro 63: Teste de K-S com nível de significância 0,05.

	Valor tabelado	Cálculo K-S	Hipótese
Agência 02	0,0418	0,054353	Rejeitada
Agência 06	0,0418	0,109342	Rejeitada
Agência 13	0,0418	0,056209	Rejeitada
Petrobrás	0,0421	0,149089	Rejeitada
Randômico	0,0418	0,060682	Rejeitada

A teste de Kolmogorov-Sminorv implica em uma distribuição uniforme dos dados, característica esta necessária para considerar uma série aleatória. Entretanto, verifica-se, na tabela acima, que todas as séries foram rejeitadas, significando, portanto, que, baseado na distribuição Kolmogorov, nenhuma delas apresenta características de aleatoriedade.

Chama atenção o fato de as agências 02 e 13 terem os valores calculados através da metodologia de Kolmogorov-Smirnov bem próximos daqueles tabelados. Estas agências ficaram, inclusive, mais próximas que a seqüência gerada pelo Matlab.

Ao se comparar os testes do qui-quadrado e de Kolmogorov-Smirnov, verifica-se que ambos indicam se o gerador de aleatórios segue a distribuição pretendida. Entretanto, existem algumas diferenças entre eles e um exemplo é o fato de o teste de Kolmogorov-Smirnov se aplicar a distribuições contínuas, enquanto que o teste do qui-quadrado se aplica a distribuições discretas (dividindo-se o conjunto em intervalos). Este último é muito sensível ao número de classes, conforme pode-se verificar no experimento anterior onde para os intervalos de 10 e 20 a série randômica foi rejeitada e para o intervalo de 50 aceita.

Outra diferença é a de que o teste do qui-quadrado necessita de grupos de dados, ao contrário do KS. Assim, a hipotética distribuição contínua do teste de

Kolmogorov-Smirnov permite examinar o ajuste para toda a amostra, tendo uma avaliação mais completa da amostra.

8.6.4. Teste de repetições de 2 tipos de observações

Quando os dados de uma série pertencem a somente 2 categorias (positivo/negativo, branco/preto, quente/frio etc) pode-se utilizar este teste a fim de se verificar a aleatoriedade da seqüência. Para se tomar uma decisão neste sentido, é necessário se contar o número de repetições nos dados e compará-lo com o número esperado de repetições.

Não foram realizados testes de repetições para os conjuntos de números randômicos e da Petrobrás pois a aplicação deste teste requer que os dados sejam dicotomizados, ou seja, pertençam a apenas 2 categorias (ex.: positivo e negativo, sim ou não etc), o que não ocorre com os conjuntos da Petrobrás e o conjunto randômico gerado pelo Matlab, que são totalmente positivos.

Quadro 64: Teste de repetições de 2 observações, com significâncias 0,05 e 0,01.

	Significância		Ag. 02	Ag. 06	Ag. 13
	0,05	0,01			
Repetições	$\pm 1,96$	$\pm 2,58$	-10,77	-4,45	-6,42

Vê-se, no quadro acima, que os valores apurados nos testes de repetições ficaram fora do intervalo para os níveis de significância 0,01 e 0,05, caracterizando, desta forma, a não aleatoriedade dos dados.

8.6.5. Teste de repetições acima e abaixo da mediana

Quando os dados em estudo são numéricos, um método comum de teste da aleatoriedade de uma seqüência focaliza quaisquer padrões que possam formar em relação à mediana amostral. Isto é, cada observação é rotulada como acima (A) ou abaixo (B) da mediana amostral e ignoram-se os valores que coincidem com a mediana (STEVENSON, 1986, p. 328-330).

Após os cálculos efetuados para as séries sob estudo, obteve-se os seguintes resultados .

Quadro 65: Teste de repetições acima e abaixo da mediana, com significâncias 0,05 e 0,01.

	Significância		Ag. 02	Ag. 06	Ag. 13	Petrobrás	Randômico
	0,05	0,01					
Mediana	+1,96	+2,58	-10,77	-6,15	-5,78	-31,4	-0,49

Os valores apurados nos testes de repetições acima e abaixo da mediana também ficaram fora do intervalo para os níveis de significância 0,01 e 0,05, com exceção da série randômica, demonstrando que, para este teste, os dados não são aleatórios.

Analisando o quadro acima, vê-se que apesar de as agências e a série da Petrobrás terem sido rejeitadas no teste, ainda assim os dados das agências ficaram bem mais próximos do limite que o da Petrobrás, ou seja, a existência de padrões nos dados das agências é inferior ao da Petrobrás.

8.6.6. Teste de repetições acima e abaixo em relação ao anterior

Por último, tem-se o teste que utiliza as repetições para cima e para baixo como medida de aleatoriedade. Mede-se cada valor da seqüência em relação ao valor anterior. Cada valor numérico é substituído por U (para cima) ou por D (para baixo), isto em relação ao valor anterior (STEVENSON, 1986, p. 330-332)..

Após os cálculos do teste obteve-se os seguintes resultados.

Quadro 66: Teste de repetições para cima e para baixo, com significâncias 0,05 e 0,01.

	Significância		Ag. 02	Ag. 06	Ag. 13	Petrobrás	Randômico
	0,05	0,01					
Cima/Baixo	+1,96	+2,58	-2,09	-7,93	-2,16	-18,78	-0,19

Das séries analisadas tanto a agência 02 como a 13 seriam aceitas como aleatórias para o nível de significância 0,01. A série randômica seria aceita para qualquer dos níveis de significância aplicados e a série da agência 06 e da Petrobrás seriam rejeitadas.

8.7. CONCLUSÕES

Este capítulo buscou verificar se o comportamento das séries poderia ser classificado como aleatório ou quase aleatório pois a simples observação delas no tempo evidencia padrões bastante irregulares. Assim, numa avaliação preliminar, as agências

parecem seguir um caminho randômico com alto grau de imprevisibilidade, sem quaisquer tendências pronunciadas, não indicando ajuste às principais formulações determinísticas.

Uma série verdadeiramente aleatória implica na inexistência de correlação entre os números, distribuição uniforme no espaço além de outras propriedades que foram descritas anteriormente. Dessa forma, a aplicação de um único teste para se avaliar o grau de aleatoriedade de uma série pode não determinar, tacitamente, a característica randômica dos dados.

Em vista do exposto, optou-se por se fazer 6 testes distintos (autocorrelação, qui-quadrado, Kolmogorov-Smirnov, além de 3 testes de repetições), de tal forma que, na hipótese de ser verificada aleatoriedade em todos os testes poderia-se considerar a série como possivelmente aleatória. Entretanto, constatou-se, através dos testes promovidos, que nenhuma das séries apresentaram características de aleatoriedade, pois todas elas, inclusive a gerada pelo Matlab, foram rejeitadas nos testes de avaliação.

As análises e testes promovidos permitem tecer alguns comentários a respeito dos dados estudados. O primeiro teste – autocorrelação -, procurou a existência ou não de ciclos entre os valores no tempo. A série randômica gerada através do aplicativo Matlab foi a que apresentou menores índices, confirmando, neste caso, a pseudo-aleatoriedade dos dados. A série da Petrobrás apresentou índices de correlação elevados, demonstrando alto grau de previsibilidade. Já as séries das agências tiveram índices significativos para considerá-las como aleatórias. Todavia, a magnitude destes coeficientes é pequena o suficiente para que se considere aceitável a suposição de previsibilidade do comportamento da série

O teste do qui-quadrado objetivou verificar o ajuste da distribuição à curva Normal. Os resultados mostraram que ao nível de significância aplicado, às distribuições das séries não se ajustam à distribuição Normal. Por este teste a hipótese de normalidade não pode ser aceita.

O terceiro teste, Kolmogorov-Smirnov, também afastou a hipótese de as distribuições empíricas aproximarem-se significativamente da Normal teórica, descartando a suposição de aleatoriedade.

Os três últimos testes (testes de corrida ou seqüência) também rejeitaram a hipótese de aleatoriedade para os níveis de significância 0,01 e 0,05 para todas as agências. Entretanto, algumas observações podem ser formuladas:

- Apesar da rejeição da hipótese, o teste de repetições para dois tipos de observações demonstrou que não se têm padrões de seqüências bem caracterizados para os valores positivos e negativos da série pois os índices calculados ficaram próximos do limite estabelecido para significância de 0,01 e 0,05;
- O teste de repetições acima e abaixo da mediana também descartou a aleatoriedade dos dados das agências mas, novamente, apresentou índices próximos do limite. Pode-se, com isto, afirmar que os conjuntos das agências não apresentam ciclos bem definidos. O índice da série da Petrobrás ficou distante dos limites, demonstrando que este conjunto de dados não é aleatório;
- Por fim, o teste de repetições acima e abaixo em relação ao anterior apresentou, para as agências 02 e 13 índices de aleatoriedade. A agência 06 teve resultado superior ao limite mas, ainda assim, próximo o que não ocorreu com a série da Petrobrás;
- A série randômica gerada pelo Matlab teve seus índices sempre dentro dos limites, caracterizando-a como aleatória para estes testes.

9. CONCLUSÃO

Neste capítulo faz-se, inicialmente, um relato sobre o problema para, na seção 2, mostrar as dificuldades existentes na criação de modelos preditivos para as séries investigadas. Finalmente, na seção 3, apresenta-se algumas sugestões de trabalhos futuros.

9.1. INTRODUÇÃO

O objetivo principal deste trabalho foi o de estudar o comportamento referente à necessidade diária de numerário de 3 (três) agências de uma instituição financeira, para fazer face aos seus compromissos com os clientes, criando um modelo computacional que respondesse a essa questão.

Procurou-se inicialmente entender o comportamento das séries, a nível de cada agência e comparativamente entre agências, com o objetivo de identificar a existência de algum padrão ou classes de padrões.

A seguir, procurou-se construir modelos lineares e não-lineares de previsões utilizando, para tanto, a aplicação de técnicas da estatística clássica (amortecimento exponencial e modelos ARIMA) e métodos de inteligência computacional (redes neurais artificiais e lógica nebulosa).

Apesar de as redes neurais artificiais terem apresentado desvios de menor magnitude quando comparados com aqueles obtidos pela aplicação dos métodos de previsão convencionais ainda assim os resultados não se mostraram satisfatórios a ponto de se poder utilizá-los na logística de distribuição de numerário.

Constatou-se que as séries são constituídas basicamente pela componente “irregularidade”, que além de ser de alta frequência apresenta um elevado índice de dispersão (variância). Em vista desta particularidade estudos sobre aleatoriedade foram realizados, onde se buscou a comprovação deste fato. Apesar de não haver sido constatada, inequivocamente, a aleatoriedade da série, a análise dos resultados mostrou que a irregularidade existente nas séries é muito forte.

Enfim, o problema se mostrou muito mais complexo do que o identificado inicialmente a despeito de todas as tentativas feitas e todos os modelos utilizados, não se conseguiu chegar a resultado que se pudesse considerar satisfatórios.

9.2. PRINCIPAIS DIFICULDADES ENCONTRADAS

O problema que se apresenta afeta todas as instituições bancárias do mundo e, ainda assim, não se encontrou material sobre o assunto. Acredita-se que em razão do segmento a que se destina – bancário, onde a confidencialidade é um dos requisitos básicos, tais trabalhos devam ser de acesso restrito. Com isto, pouca literatura discorrendo sobre o problema foi localizada, dificultando comparações e avaliações com outras pesquisas e métodos já testados.

A base de dados disponível também se mostrou pequena para o trabalho. Em vista dos problemas comentados no texto desta dissertação (grande oscilação dos dados, irregularidades nas séries, etc), acredita-se que uma série mais longa de dados possibilitasse uma maior chance de sucesso na modelagem de problema. Entretanto, só existem registros disponíveis a partir de 1998, pois os dados anteriores, na sua maioria, não estão em mídia magnética e/ou não são confiáveis para se trabalhar.

A inexistência de outras variáveis que possuam correlação com a movimentação das agências, tais como: número de clientes, perfil de renda dos clientes, quantidade de contas de comerciantes, indústrias atendidas pela agência, datas e formas de pagamentos de salários das contas de pessoas jurídicas, sazonalidades, feiras e congressos na região, safras agrícolas etc. poderiam auxiliar no mapeamento das características das séries. Entretanto, estas informações não estão disponíveis e seria necessária uma pesquisa de campo para se obtê-las, o que não foi possível fazer durante o período de desenvolvimento desta dissertação.

O dinamismo que afeta todo o sistema bancário também é fator que necessitaria ser investigado. A criação de novas agências ou a inauguração de outros bancos nas praças sob estudo afeta todo o sistema local. Portanto, haveria necessidade de se verificar se ocorreram mudanças deste tipo na localidade, ou seja, se novos bancos começaram a operar no local ou se houve fechamento ou abertura de agências.

O período sob avaliação também foi caracterizado por um grande salto tecnológico no Banco objeto de estudo. Houve a implantação de vários terminais de

auto-atendimento por todo o País e a sua criação significa redução no movimento da agência pois o dinheiro transacionado nestes terminais não é contabilizado na agência. Dessa forma, variações significativas no comportamento podem ocorrer. Buscou-se descobrir se tal mudança – implantação de máquinas de auto-atendimento, ocorreram nas localidades mas este tipo de informação não estava disponível.

Várias foram as estratégias utilizadas para a definição do intervalo ideal que melhor representasse a performance de acerto. Foram realizados estudos com o uso de intervalos de frequência, quantização dos dados etc a fim de se descobrir qual a melhor métrica para se aplicar aos registros de saída, visando a melhoria na performance de previsão. Entretanto, não se conseguiu um balanceamento ideal entre a quantidade de intervalos e os índices de acerto, ou seja, para poucos intervalos a performance era boa mas ao aumentar-se o número de intervalos os resultados caíam significativamente.

O tratamento dos *outliers* também não apresentou resultados satisfatórios. Não se detectou regularidades nem padrões nestes dados. Os registros, nestas situações, apesar de parecerem espúrios em função da aparente divergência com os demais, foram conferidos e checados com a fonte e correspondem efetivamente ao movimento. Há, portanto, necessidade de se aprofundar nas causas que levaram a se ter dados com estas características.

9.3. MODELOS APLICADOS

9.3.1. Modelo de amortecimento exponencial e ARIMA

Os estudos relativos aos modelos de amortecimento exponencial e ARIMA foram realizados com 1.059 registros e os valores acima de 2DP foram trazidos para este patamar. Desta massa de dados foi aplicada a fórmula sobre os 1.009 registros iniciais e procurou-se prever os 50 registros finais. Cabe ressaltar que a construção do modelo foi feita com os primeiros 1.009 registros e cada previsão era feita com todos os registros anteriores ao corrente, ou seja, para cálculo do 1.010º. utilizou-se o modelo com 1.009 registros. Já o cálculo do 1.011º. foi feito com os 1.010 registros anteriores e assim sucessivamente.

9.3.2. Modelo Neural

A construção do modelo neural foi a que consumiu a maior parte do tempo de pesquisa em função do fato de ser considerado, atualmente, o paradigma mais poderoso para se aproximar uma função. Assim, várias redes foram construídas durante a fase de pesquisa bem como muitos parâmetros de ajuste – taxas de treinamento, funções de aprendizagem etc. -, foram aplicados.

9.3.3. Modelo Nebuloso

No modelo nebuloso utilizou-se 1.059 registros. Desse total foram usados em treinamento 750 e o restante, ou seja, os 309 últimos registros, em teste.

Para estes experimentos utilizou-se somente o tipo de *fuzzificação* 4 pois foi neste tipo de *fuzzificação* que se obteve as melhores performances na fase de treinamento.

9.3.4. Comparações Entre os Diversos Modelos

A título de comparação, segue abaixo quadro comparativo dos diversos modelos aplicados na solução do problema. Cabe ressaltar que o quadro só apresenta os melhores resultados.

Quadro 67: Quadro comparativo dos modelos utilizados.

	Amortecimento exponencial	Arima	Redes Neurais	Lógica Nebulosa
Agência 02	20,0	26,0	23,0	11,4
Agência 06	22,0	20,0	29,6	8,9
Agência 13	2,0	6,0	21,1	11,0

9.4. TRABALHOS FUTUROS

Em termos de trabalhos futuros, muito ainda há a se fazer pois o problema se mostrou muito mais complexo do que se imaginava inicialmente. Abaixo têm-se as principais linhas a serem abordadas em estudos futuros.

- Tratamento dos dados – levantar e avaliar a aplicação de novas formas de transformação e normalização dos dados de maneira a melhor evidenciar os possíveis padrões existentes;

- Análise multivariável – tentar identificar e obter outras variáveis dentro do contexto do problema que possam ajudar na explicação do fenômeno;
- Modelos preditivos – conceber e avaliar o desempenho de um modelo híbrido com base na rede Elman ou mesmo da rede TDNN e regras nebulosas;
- Métricas de desempenho – conceber uma métrica para avaliar o desempenho dos modelos, que não penalize de forma desigual valores elevados e valores baixos e que melhor reflita a realidade do problema e a expectativa do usuário.

Anexo

1. ANÁLISE DOS OUTLIERS

Quadro 1: Ocorrências verificadas nas agências 02, 06 e 13 em função dos feriados do ano de 1998.

	Dia	Semana	Mês	Útil	Valor	Agência	Observações	
Ano de 1998	05/01/1998	2	5	2	2126,00	Ag. 06	Segundo dia útil do ano.	FÉRIAS
	06/01/1998	3	6	3	2905,64	Ag. 06	Terceiro dia útil do ano.	
	13/01/1998	3	13	8	2090,93	Ag. 06	Sem justificativa.	
	16/01/1998	6	16	11	1645,40	Ag. 06	Sem justificativa.	
	21/01/1998	4	21	14	2259,71	Ag. 06	Sem justificativa.	
	22/01/1998	5	22	15	-464,82	Ag. 02	Sem justificativa.	
	26/01/1998	2	26	17	1940,78	Ag. 06	Sem justificativa.	
	27/01/1998	3	27	18	2300,17	Ag. 06	Sem justificativa.	
	03/02/1998	3	3	2	1466,19	Ag. 06	Sem justificativa.	
	05/02/1998	5	5	4	-115,33	Ag. 13	Sem justificativa.	
	10/02/1998	3	10	7	1867,25	Ag. 06	Sem justificativa.	
	13/02/1998	6	13	10	-1090,00	Ag. 06	Sem justificativa.	
	17/02/1998	3	17	12	2050,40	Ag. 06	Sem justificativa.	
	25/02/1998	4	25	16	1035,59	Ag. 02	Quarta-feira de cinzas.	
	25/02/1998	4	25	16	2355,93	Ag. 06	Quarta-feira de cinzas.	
	26/02/1998	5	26	17	4227,61	Ag. 06	Quinta-feira após carnaval.	
	27/02/1998	6	27	18	1626,01	Ag. 06	Sexta-feira após carnaval.	
	02/03/1998	2	2	1	1681,45	Ag. 06	Sem justificativa.	
	03/03/1998	3	3	2	1841,36	Ag. 06	Sem justificativa.	
	13/03/1998	6	13	10	-558,56	Ag. 02	Sem justificativa.	
	16/03/1998	2	16	11	811,84	Ag. 02	Sem justificativa.	
	02/04/1998	5	2	2	-128,69	Ag. 13	Sem justificativa.	
	13/04/1998	2	13	7	438,21	Ag. 02	Após semana santa.	
	13/04/1998	2	13	7	173,53	Ag. 13	Após semana santa.	
	14/04/1998	3	14	8	2170,93	Ag. 06	Após semana santa.	
	30/04/1998	5	30	19	-535,00	Ag. 02	Véspera de feriado. (01/05)	
	30/04/1998	5	30	19	-157,13	Ag. 13	Véspera de feriado. (01/05)	
	18/05/1998	2	18	11	365,19	Ag. 02	Sem justificativa.	
	19/05/1998	3	19	12	354,59	Ag. 02	Sem justificativa.	
	02/06/1998	3	2	2	-135,07	Ag. 13	Sem justificativa.	
	12/06/1998	6	12	9	-590,20	Ag. 02	Corpus Christi.	
	15/06/1998	2	15	10	748,43	Ag. 02	Após Corpus Christi.	
26/06/1998	6	26	19	-494,10	Ag. 02	Sem justificativa.		
02/07/1998	5	2	2	-133,09	Ag. 13	Sem justificativa.		
08/09/1998	3	8	5	366,92	Ag. 02	Após feriado. (07/09)		
25/09/1998	6	25	18	-372,39	Ag. 02	Sem justificativa.		
05/10/1998	2	5	3	-159,67	Ag. 13	Sem justificativa.		
13/10/1998	3	13	8	397,98	Ag. 02	Após feriado. (12/10)		
19/10/1998	2	19	12	384,68	Ag. 02	Sem justificativa.		
26/10/1998	2	26	17	385,71	Ag. 02	Sem justificativa.		
04/11/1998	4	4	2	-128,14	Ag. 13	Após feriado. (02/11)		
02/12/1998	4	2	2	-152,39	Ag. 13	Sem justificativa.		
14/12/1998	2	14	10	380,35	Ag. 02	Sem justificativa.		
18/12/1998	6	18	14	-130,14	Ag. 13	Sem justificativa.		
28/12/1998	2	28	19	175,84	Ag. 13	Após feriado. (25/12)		

Quadro 2: Ocorrências verificadas nas agências 02, 06 e 13 em função dos feriados do ano de 1999.

	Dia	Sem.	Mês	Útil	Valor	Agência	Observações	
Ano de 1999	04/01/1999	2	4	1	484,18	Ag. 02	Primeiro dia útil do ano.	FÉRIAS
	05/01/1999	3	5	2	1537,24	Ag. 06	Sem justificativa.	
	06/01/1999	4	6	3	2545,79	Ag. 06	Sem justificativa.	
	07/01/1999	5	7	4	-486,21	Ag. 02	Sem justificativa.	
	08/01/1999	6	8	5	-145,69	Ag. 13	Sem justificativa.	
	19/01/1999	3	19	12	1473,33	Ag. 06	Sem justificativa.	
	26/01/1999	3	26	17	1467,04	Ag. 06	Sem justificativa.	
	28/01/1999	5	28	19	-500,00	Ag. 02	Sem justificativa.	
	03/02/1999	4	3	3	-133,87	Ag. 13	Sem justificativa.	
	12/02/1999	6	12	10	-1065,92	Ag. 06	Véspera de carnaval.	
	17/02/1999	4	17	11	693,39	Ag. 02	Quarta-feira de cinzas.	
	17/02/1999	4	17	11	1472,35	Ag. 06	Quarta-feira de cinzas.	
	18/02/1999	5	18	12	3183,85	Ag. 06	Quinta-feira após carnaval.	
	22/02/1999	2	22	14	1955,81	Ag. 06	Sem justificativa.	
	23/02/1999	3	23	15	1654,47	Ag. 06	Sem justificativa.	
	26/02/1999	6	26	18	-399,27	Ag. 02	Sem justificativa.	
	31/03/1999	4	31	23	-145,61	Ag. 13	Véspera semana Santa.	
	28/05/1999	6	28	20	-125,06	Ag. 13	Sem justificativa.	
	29/06/1999	3	29	20	-118,99	Ag. 13	Véspera S.Pedro (30/06)	
	02/07/1999	6	2	2	-433,60	Ag. 02	Sem justificativa.	
	30/07/1999	6	30	22	-247,34	Ag. 13	Sem justificativa.	
	03/08/1999	3	3	2	-442,25	Ag. 02	Sem justificativa.	
	30/08/1999	2	30	21	-124,42	Ag. 13	Sem justificativa.	
	02/09/1999	5	2	2	-458,66	Ag. 02	Sem justificativa.	
	03/09/1999	6	3	3	-477,70	Ag. 02	Sem justificativa.	
	28/10/1999	5	28	19	-136,80	Ag. 13	Sem justificativa.	
	04/11/1999	5	4	3	-395,23	Ag. 02	Após feriado. (02/11)	
26/11/1999	6	26	18	-128,86	Ag. 13	Sem justificativa.		
22/12/1999	4	22	16	-248,10	Ag. 13	Sem justificativa.		
27/12/1999	2	27	19	195,71	Ag. 13	Após feriado.		
29/12/1999	4	29	21	-210,15	Ag. 13	Antevéspera feriado. (31/12)		

Quadro 3: Ocorrências verificadas nas agências 02, 06 e 13 em função dos feriados do ano de 2000 e 2001, existentes até 31/03.

	Dia	Sem.	Mês	Útil	Valor	Agência	Observações	
Ano de 2000	03/01/2000	2	3	1	520,52	Ag. 02	Primeiro dia útil do ano.	FÉRIAS
	03/01/2000	2	3	1	176,05	Ag. 13	Primeiro dia útil do ano.	
	07/01/2000	6	7	5	-135,06	Ag. 13	Sem justificativa.	
	19/01/2000	4	19	13	1539,49	Ag. 06	Sem justificativa.	
	25/02/2000	6	25	19	-127,83	Ag. 13	Sem justificativa.	
	02/03/2000	5	2	2	-389,62	Ag. 02	Antevéspera de carnaval.	
	02/03/2000	5	2	2	-1448,85	Ag. 06	Antevéspera de carnaval.	
	03/03/2000	6	3	3	-384,23	Ag. 02	Véspera de carnaval.	
	03/03/2000	6	3	3	-1470,37	Ag. 06	Véspera de carnaval.	
	08/03/2000	4	8	4	598,65	Ag. 02	Quarta-feira de cinzas.	
	09/03/2000	5	9	5	2952,06	Ag. 06	Quinta-feira após carnaval.	
	13/03/2000	2	13	7	1438,41	Ag. 06	Sem justificativa.	
	28/03/2000	3	28	18	-130,72	Ag. 13	Sem justificativa.	
	04/04/2000	3	4	2	-479,67	Ag. 02	Sem justificativa.	
	20/04/2000	5	20	14	-905,65	Ag. 06	Véspera Semana Santa	
	20/04/2000	5	20	14	-221,20	Ag. 13	Véspera Semana Santa	
	28/04/2000	6	28	19	-922,49	Ag. 06	Sem justificativa.	
	26/05/2000	6	26	19	-159,41	Ag. 13	Sem justificativa.	
	02/06/2000	6	2	2	-371,20	Ag. 02	Sem justificativa.	
	28/06/2000	4	28	19	-228,56	Ag. 13	Sem justificativa.	
	03/07/2000	2	3	1	522,84	Ag. 02	Sem justificativa.	
	02/08/2000	4	2	2	-395,36	Ag. 02	Sem justificativa.	
	15/08/2000	3	15	11	956,67	Ag. 02	Sem justificativa.	
	11/09/2000	2	11	6	-178,96	Ag. 13	Após feriado. (07/09)	
	12/09/2000	3	12	7	1980,57	Ag. 06	Após feriado. (07/09)	
	15/09/2000	6	15	10	360,07	Ag. 02	Sem justificativa.	
	19/09/2000	3	19	12	1732,31	Ag. 06	Sem justificativa.	
	27/09/2000	4	27	18	-140,57	Ag. 13	Sem justificativa.	
	18/10/2000	4	18	12	2258,38	Ag. 06	Sem justificativa.	
	24/10/2000	3	24	16	1758,58	Ag. 06	Sem justificativa.	
	03/11/2000	6	3	2	-394,41	Ag. 02	Após feriado. (02/11)	
	07/11/2000	3	7	4	1469,68	Ag. 06	Sem justificativa.	
13/11/2000	2	13	8	966,67	Ag. 02	Antevéspera feriado. (15/11)		
21/11/2000	3	21	13	2080,20	Ag. 06	Sem justificativa.		
28/11/2000	3	28	18	-113,32	Ag. 13	Sem justificativa.		
01/12/2000	6	1	1	919,40	Ag. 13	Sem justificativa.		
21/12/2000	5	21	15	-265,84	Ag. 13	Antevéspera feriado. (25/12)		
22/12/2000	6	22	16	-143,79	Ag. 13	Véspera de feriado. (25/12)		
26/12/2000	3	26	17	214,11	Ag. 13	Após feriado. (25/12)		
28/12/2000	5	28	19	-1370,65	Ag. 06	Antevéspera feriado. (01/01)		
Ano de 2001	02/01/2001	3	2	1	409,73	Ag. 02	Primeiro dia útil do ano.	FÉRIAS
	02/01/2001	3	2	1	158,99	Ag. 13	Primeiro dia útil do ano.	
	26/01/2001	6	26	19	-138,46	Ag. 13	Sem justificativa.	
	06/02/2001	3	6	4	1776,16	Ag. 06	Sem justificativa.	
	12/02/2001	2	12	8	361,26	Ag. 02	Sem justificativa.	
	23/02/2001	6	23	17	-1419,82	Ag. 06	Véspera Carnaval	
	28/02/2001	4	28	18	618,99	Ag. 02	Quarta-feira de cinzas.	
	28/02/2001	4	28	18	154,61	Ag. 13	Quarta-feira de cinzas.	

2. TRATAMENTO DOS *OUTLIERS*

Quadro 4: Agência 02 – Distribuição mensal dos outliers.

Mês	Quantidade	%
Janeiro	6	14,29
Fevereiro	5	11,90
Março	5	11,90
Abril	3	7,14
Mai	2	4,76
Junho	4	9,52
Julho	2	4,76
Agosto	3	7,14
Setembro	5	11,90
Outubro	3	7,14
Novembro	3	7,14
Dezembro	1	2,38
Totais	42	100,00

Quadro 5: Agência 06 – Distribuição mensal dos outliers.

Mês	Quantidade	%
Janeiro	12	28,57
Fevereiro	14	33,33
Março	6	14,29
Abril	3	7,14
Mai	0	0,00
Junho	0	0,00
Julho	0	0,00
Agosto	0	0,00
Setembro	2	4,76
Outubro	2	4,76
Novembro	2	4,76
Dezembro	1	2,38
Totais	42	100,00

Quadro 6: Agência 13 – Distribuição mensal dos outliers.

Mês	Quantidade	%
Janeiro	5	12,50
Fevereiro	4	10,00
Março	2	5,00
Abril	4	10,00
Mai	2	5,00
Junho	3	7,50
Julho	2	5,00
Agosto	1	2,50
Setembro	2	5,00
Outubro	2	5,00
Novembro	3	7,50
Dezembro	10	25,00
Totais	40	100,00

Quadro 7: Agência 02 – outliers classificados por dia útil seguido da data. A coluna “semana” representa o dia da semana, segundo a seguinte correlação: 2 corresponde à segunda-feira, 3 à terça-feira e assim sucessivamente. Na coluna “mês” temos os dias do mês e, por último, na coluna “útil” o dia útil a que se refere.

Agência 02 - Outliers classificados por dia útil seguido da data									
Data	Semana	Mês	Útil	Valor	Data	Semana	Mês	Útil	Valor
04/01/1999	2	4	1	484,18	12/02/2001	2	12	8	361,26
03/01/2000	2	3	1	520,52	12/06/1998	6	12	9	-590,20
03/07/2000	2	3	1	522,84	13/03/1998	6	13	10	-558,56
02/01/2001	3	2	1	409,73	15/06/1998	2	15	10	748,43
02/07/1999	6	2	2	-433,60	14/12/1998	2	14	10	380,35
03/08/1999	3	3	2	-442,25	15/09/2000	6	15	10	360,07
02/09/1999	5	2	2	-458,66	16/03/1998	2	16	11	811,84
02/03/2000	5	2	2	-389,62	18/05/1998	2	18	11	365,19
04/04/2000	3	4	2	-479,67	17/02/1999	4	17	11	693,39
02/06/2000	6	2	2	-371,20	15/08/2000	3	15	11	956,67
02/08/2000	4	2	2	-395,36	19/05/1998	3	19	12	354,59
03/11/2000	6	3	2	-394,41	19/10/1998	2	19	12	384,68
03/09/1999	6	3	3	-477,70	22/01/1998	5	22	15	-464,82
04/11/1999	5	4	3	-395,23	25/02/1998	4	25	16	1035,59
03/03/2000	6	3	3	-384,23	26/10/1998	2	26	17	385,71
07/01/1999	5	7	4	-486,21	25/09/1998	6	25	18	-372,39
08/03/2000	4	8	4	598,65	26/02/1999	6	26	18	-399,27
08/09/1998	3	8	5	366,92	28/02/2001	4	28	18	618,99
13/04/1998	2	13	7	438,21	30/04/1998	5	30	19	-535,00
13/10/1998	3	13	8	397,98	26/06/1998	6	26	19	-494,10
13/11/2000	2	13	8	966,67	28/01/1999	5	28	19	-500,00

Quadro 8: Agência 06 – Outliers classificados por dia útil seguido da data. A coluna “semana” representa o dia da semana, segundo a seguinte correlação: 2 corresponde à segunda-feira, 3 à terça-feira e assim sucessivamente. Na coluna “mês” temos os dias do mês e, por último, na coluna “útil” o dia útil a que se refere.

Agência 06 - Outliers classificados por dia útil seguido da data									
Data	Semana	Mês	Útil	Valor	Data	Semana	Mês	Útil	Valor
02/03/1998	2	2	1	1681,45	17/02/1998	3	17	12	2050,40
05/01/1998	2	5	2	2126,00	19/01/1999	3	19	12	1473,33
03/02/1998	3	3	2	1466,19	18/02/1999	5	18	12	3183,85
03/03/1998	3	3	2	1841,36	19/09/2000	3	19	12	1732,31
05/01/1999	3	5	2	1537,24	18/10/2000	4	18	12	2258,38
02/03/2000	5	2	2	-1448,85	19/01/2000	4	19	13	1539,49
06/01/1998	3	6	3	2905,64	21/11/2000	3	21	13	2080,20
06/01/1999	4	6	3	2545,79	21/01/1998	4	21	14	2259,71
03/03/2000	6	3	3	-1470,37	22/02/1999	2	22	14	1955,81
07/11/2000	3	7	4	1469,68	20/04/2000	5	20	14	-905,65
06/02/2001	3	6	4	1776,16	23/02/1999	3	23	15	1654,47
09/03/2000	5	9	5	2952,06	25/02/1998	4	25	16	2355,93
10/02/1998	3	10	7	1867,25	24/10/2000	3	24	16	1758,58
13/03/2000	2	13	7	1438,41	26/01/1998	2	26	17	1940,78
12/09/2000	3	12	7	1980,57	26/02/1998	5	26	17	4227,61
13/01/1998	3	13	8	2090,93	26/01/1999	3	26	17	1467,04
14/04/1998	3	14	8	2170,93	23/02/2001	6	23	17	-1419,82
13/02/1998	6	13	10	-1090,00	27/01/1998	3	27	18	2300,17
12/02/1999	6	12	10	-1065,92	27/02/1998	6	27	18	1626,01
16/01/1998	6	16	11	1645,40	28/04/2000	6	28	19	-922,49
17/02/1999	4	17	11	1472,35	28/12/2000	5	28	19	-1370,65

Quadro 9: Agência 13 – Outliers classificados por dia útil seguido da data. A coluna “semana” representa o dia da semana, segundo a seguinte correlação: 2 corresponde à segunda-feira, 3 à terça-feira e assim sucessivamente. Na coluna “mês” temos os dias do mês e, por último, na coluna “útil” o dia útil a que se refere.

Agência 13 - Outliers classificados por dia útil seguido da data									
Data	Semana	Mês	Útil	Valor	Data	Semana	Mês	Útil	Valor
03/01/2000	2	3	1	176,05	26/12/2000	3	26	17	214,11
01/12/2000	6	1	1	919,40	26/11/1999	6	26	18	-128,86
02/01/2001	3	2	1	158,99	28/03/2000	3	28	18	-130,72
02/04/1998	5	2	2	-128,69	27/09/2000	4	27	18	-140,57
02/06/1998	3	2	2	-135,07	28/11/2000	3	28	18	-113,32
02/07/1998	5	2	2	-133,09	28/02/2001	4	28	18	154,61
04/11/1998	4	4	2	-128,14	30/04/1998	5	30	19	-157,13
02/12/1998	4	2	2	-152,39	28/12/1998	2	28	19	175,84
05/10/1998	2	5	3	-159,67	28/10/1999	5	28	19	-136,80
03/02/1999	4	3	3	-133,87	27/12/1999	2	27	19	195,71
05/02/1998	5	5	4	-115,33	25/02/2000	6	25	19	-127,83
08/01/1999	6	8	5	-145,69	26/05/2000	6	26	19	-159,41
07/01/2000	6	7	5	-135,06	28/06/2000	4	28	19	-228,56
11/09/2000	2	11	6	-178,96	26/01/2001	6	26	19	-138,46
13/04/1998	2	13	7	173,53	28/05/1999	6	28	20	-125,06
18/12/1998	6	18	14	-130,14	29/06/1999	3	29	20	-118,99
20/04/2000	5	20	14	-221,20	30/08/1999	2	30	21	-124,42
21/12/2000	5	21	15	-265,84	29/12/1999	4	29	21	-210,15
22/12/1999	4	22	16	-248,10	30/07/1999	6	30	22	-247,34
22/12/2000	6	22	16	-143,79	31/03/1999	4	31	23	-145,61

Quadro 10: Ag. 02 – Outliers classificados por dia da semana. A coluna “semana” representa o dia da semana, segundo a seguinte correlação: 2 corresponde à segunda-feira, 3 à terça-feira e assim sucessivamente. Na coluna “mês” têm-se os dias do mês e, por último, na coluna “útil” o a que se refere.

Agência 02 - Outliers classificados por dia da semana seguido da data									
Data	Semana	Mês	Útil	Valor	Data	Semana	Mês	Útil	Valor
16/03/1998	2	16	11	811,84	08/03/2000	4	8	4	598,65
13/04/1998	2	13	7	438,21	02/08/2000	4	2	2	-395,36
18/05/1998	2	18	11	365,19	28/02/2001	4	28	18	618,99
15/06/1998	2	15	10	748,43	22/01/1998	5	22	15	-464,82
19/10/1998	2	19	12	384,68	30/04/1998	5	30	19	-535,00
26/10/1998	2	26	17	385,71	07/01/1999	5	7	4	-486,21
14/12/1998	2	14	10	380,35	28/01/1999	5	28	19	-500,00
04/01/1999	2	4	1	484,18	02/09/1999	5	2	2	-458,66
03/01/2000	2	3	1	520,52	04/11/1999	5	4	3	-395,23
03/07/2000	2	3	1	522,84	02/03/2000	5	2	2	-389,62
13/11/2000	2	13	8	966,67	13/03/1998	6	13	10	-558,56
12/02/2001	2	12	8	361,26	12/06/1998	6	12	9	-590,20
19/05/1998	3	19	12	354,59	26/06/1998	6	26	19	-494,10
08/09/1998	3	8	5	366,92	25/09/1998	6	25	18	-372,39
13/10/1998	3	13	8	397,98	26/02/1999	6	26	18	-399,27
03/08/1999	3	3	2	-442,25	02/07/1999	6	2	2	-433,60
04/04/2000	3	4	2	-479,67	03/09/1999	6	3	3	-477,70
15/08/2000	3	15	11	956,67	03/03/2000	6	3	3	-384,23
02/01/2001	3	2	1	409,73	02/06/2000	6	2	2	-371,20
25/02/1998	4	25	16	1035,59	15/09/2000	6	15	10	360,07
17/02/1999	4	17	11	693,39	03/11/2000	6	3	2	-394,41

Quadro 11: Agência 06 – Outliers classificados por dia da semana seguido da data. A coluna “semana” representa o dia da semana, segundo a seguinte correlação: 2 corresponde à segunda-feira, 3 à terça-feira e assim sucessivamente. Na coluna “mês” têm-se os dias do mês e, por último, na coluna “útil” o a que se refere.

Agência 06 - Outliers classificados por dia da semana seguido da data									
Data	Semana	Mês	Útil	Valor	Data	Semana	Mês	Útil	Valor
05/01/1998	2	5	2	2126,00	21/11/2000	3	21	13	2080,20
26/01/1998	2	26	17	1940,78	06/02/2001	3	6	4	1776,16
02/03/1998	2	2	1	1681,45	21/01/1998	4	21	14	2259,71
22/02/1999	2	22	14	1955,81	25/02/1998	4	25	16	2355,93
13/03/2000	2	13	7	1438,41	06/01/1999	4	6	3	2545,79
06/01/1998	3	6	3	2905,64	17/02/1999	4	17	11	1472,35
13/01/1998	3	13	8	2090,93	19/01/2000	4	19	13	1539,49
27/01/1998	3	27	18	2300,17	18/10/2000	4	18	12	2258,38
03/02/1998	3	3	2	1466,19	26/02/1998	5	26	17	4227,61
10/02/1998	3	10	7	1867,25	18/02/1999	5	18	12	3183,85
17/02/1998	3	17	12	2050,40	02/03/2000	5	2	2	-1448,85
03/03/1998	3	3	2	1841,36	09/03/2000	5	9	5	2952,06
14/04/1998	3	14	8	2170,93	20/04/2000	5	20	14	-905,65
05/01/1999	3	5	2	1537,24	28/12/2000	5	28	19	-1370,65
19/01/1999	3	19	12	1473,33	16/01/1998	6	16	11	1645,40
26/01/1999	3	26	17	1467,04	13/02/1998	6	13	10	-1090,00
23/02/1999	3	23	15	1654,47	27/02/1998	6	27	18	1626,01
12/09/2000	3	12	7	1980,57	12/02/1999	6	12	10	-1065,92
19/09/2000	3	19	12	1732,31	03/03/2000	6	3	3	-1470,37
24/10/2000	3	24	16	1758,58	28/04/2000	6	28	19	-922,49
07/11/2000	3	7	4	1469,68	23/02/2001	6	23	17	-1419,82

Quadro 12: Agência 13 – Outliers classificados por dia da semana seguido da data. A coluna “semana” representa o dia da semana, segundo a seguinte correlação: 2 corresponde à segunda-feira, 3 à terça-feira e assim sucessivamente. Na coluna “mês” têm-se os dias do mês e, por último, na coluna “útil” o a que se refere.

Agência 13 - Outliers classificados por dia da semana seguido da data									
Data	Semana	Mês	Útil	Valor	Data	Semana	Mês	Útil	Valor
13/04/1998	2	13	7	173,53	27/09/2000	4	27	18	-140,57
05/10/1998	2	5	3	-159,67	28/02/2001	4	28	18	154,61
28/12/1998	2	28	19	175,84	05/02/1998	5	5	4	-115,33
30/08/1999	2	30	21	-124,42	02/04/1998	5	2	2	-128,69
27/12/1999	2	27	19	195,71	30/04/1998	5	30	19	-157,13
03/01/2000	2	3	1	176,05	02/07/1998	5	2	2	-133,09
11/09/2000	2	11	6	-178,96	28/10/1999	5	28	19	-136,80
02/06/1998	3	2	2	-135,07	20/04/2000	5	20	14	-221,20
29/06/1999	3	29	20	-118,99	21/12/2000	5	21	15	-265,84
28/03/2000	3	28	18	-130,72	18/12/1998	6	18	14	-130,14
28/11/2000	3	28	18	-113,32	08/01/1999	6	8	5	-145,69
26/12/2000	3	26	17	214,11	28/05/1999	6	28	20	-125,06
02/01/2001	3	2	1	158,99	30/07/1999	6	30	22	-247,34
04/11/1998	4	4	2	-128,14	26/11/1999	6	26	18	-128,86
02/12/1998	4	2	2	-152,39	07/01/2000	6	7	5	-135,06
03/02/1999	4	3	3	-133,87	25/02/2000	6	25	19	-127,83
31/03/1999	4	31	23	-145,61	26/05/2000	6	26	19	-159,41
22/12/1999	4	22	16	-248,10	01/12/2000	6	1	1	919,40
29/12/1999	4	29	21	-210,15	22/12/2000	6	22	16	-143,79
28/06/2000	4	28	19	-228,56	26/01/2001	6	26	19	-138,46

3. MODELOS ESTATÍSTICOS UTILIZADOS

Quadro 13: Previsão da série da agência 02 (sem outliers) segundo os modelos de Box-Jenkins e amortecimento exponencial.

Data	Agência 02 - Box-Jenkins				Agência 02 - Amortecimento exponencial			
	Real	Previsto	Dif. %	Acertos	Real	Previsto	Dif. %	Acertos
16/01/02	147,03	174,53	18,7	X	147,03	265,28	80,43	
17/01/02	126,53	132,29	4,55	X	126,53	151,6	19,81	X
18/01/02	-62,11	78,04	-225,65		-62,11	102,76	-265,45	
21/01/02	327,06	151,53	-53,67		327,06	197,44	-39,63	
22/01/02	138,47	99,54	-28,11		138,47	109,79	-20,71	
23/01/02	132,25	145,77	10,22	X	132,25	238,93	80,67	
24/01/02	88,83	123,82	39,39		88,83	146,2	64,58	
25/01/02	187,76	10,39	-94,47		187,76	65,9	-64,9	
28/01/02	354,12	203,28	-42,6		354,12	226,36	-36,08	
29/01/02	170,84	118,52	-30,63		170,84	116,23	-31,97	
30/01/02	85,14	129,68	52,31		85,14	214,99	152,51	
31/01/02	107,84	78,78	-26,95		107,84	133,37	23,67	
01/02/02	67,74	66,64	-1,62	X	67,74	93,01	37,3	
04/02/02	189,65	243,46	28,37		189,65	256,12	35,05	
05/02/02	56,01	104,18	86,0		56,01	128,43	129,3	
06/02/02	0	20,98	X		0	185,95	0	
07/02/02	185,28	54,07	-70,82		185,28	127,59	-31,14	
08/02/02	27,01	61,79	128,77		27,01	87,29	223,18	
13/02/02	372,61	181,24	-51,36		372,61	241,08	-35,3	
14/02/02	198,96	100,55	-49,46		198,96	112,26	-43,58	
15/02/02	173,64	62,95	-63,75		173,64	143,02	-17,63	X
18/02/02	372,61	119,78	-67,85		372,61	140,69	-62,24	
19/02/02	135,07	78,01	-42,24		135,07	73,51	-45,58	
20/02/02	122,47	254,11	107,49		122,47	271,37	121,58	
21/02/02	131,96	101,86	-22,81		131,96	131,86	-0,08	X
22/02/02	121,32	62,28	-48,66		121,32	150,34	23,92	
25/02/02	263,72	159,73	-39,43		263,72	193,13	-26,77	
26/02/02	60,2	113,83	89,09		60,2	87,55	45,43	
27/02/02	90,17	126,01	39,75		90,17	237,27	163,14	
28/02/02	82,06	78,74	-4,05	X	82,06	131,89	60,72	
01/03/02	50,1	58,27	16,31	X	50,1	143,77	186,97	
04/03/02	190,54	149,33	-21,63		190,54	209,11	9,75	X
05/03/02	50,72	49,85	-1,72	X	50,72	81,36	60,41	
06/03/02	47,82	60,77	27,08		47,82	203,98	326,56	
07/03/02	49,48	47,71	-3,58	X	49,48	120,61	143,76	
08/03/02	67,53	42,21	-37,49		67,53	122,58	81,52	
11/03/02	372,61	151,43	-59,36		372,61	204,85	-45,02	
12/03/02	120,74	73,2	-39,37		120,74	74,32	-38,45	
13/03/02	129,54	106,99	-17,41	X	129,54	168,29	29,91	
14/03/02	88,53	80,99	-8,52	X	88,53	104,4	17,93	X
15/03/02	50,97	85,87	68,47		50,97	109,68	115,19	
18/03/02	256,68	222,46	-13,33	X	256,68	244,12	-4,89	X
19/03/02	87,38	77,7	-11,08	X	87,38	84,91	-2,83	X
20/03/02	131,8	48,1	-63,51		131,8	159	20,64	
21/03/02	104,62	67,17	-35,8		104,62	100,63	-3,81	X
22/03/02	89,51	65,92	-26,35		89,51	95,72	6,94	X
25/03/02	308,03	175,91	-42,89		308,03	247,02	-19,81	X
26/03/02	62,47	76,28	22,11		62,47	85,48	36,83	
27/03/02	93,46	79,57	-14,86	X	93,46	152,68	63,36	
28/03/02	79,59	58,78	-26,15		79,59	101,53	27,57	
			Total Ac.	26%			Total Ac.	20%

Quadro 14: Previsão da série da agência 06 (sem outliers) segundo os modelos de Box-Jenkins e amortecimento exponencial.

Data	Agência 06 - Box-Jenkins				Agência 06 - Amortecimento exponencial			
	Real	Previsto	Dif. %	Acertos	Real	Previsto	Dif. %	Acertos
16/01/02	283,27	350,73	23,81		283,27	500,34	76,63	
17/01/02	232,54	191,98	-17,44	X	232,54	231,77	-0,33	X
18/01/02	152,03	204,84	34,74		152,03	180,68	18,84	X
21/01/02	602,29	386,12	-35,89		602,29	405,24	-32,72	
22/01/02	237,28	320,27	34,98		237,28	230,7	-2,77	X
23/01/02	275,68	261,56	-5,12	X	275,68	414,4	50,32	
24/01/02	60,14	224,35	273,05		60,14	233,12	287,63	
25/01/02	-68,97	141,11	-304,6		-68,97	163,16	-336,57	
28/01/02	435,53	375,94	-13,68	X	435,53	466,67	7,15	X
29/01/02	31,07	256,72	726,26		31,07	208,44	570,87	
30/01/02	167,5	204,03	21,81	X	167,5	327,79	95,7	
31/01/02	26,04	147,49	466,4		26,04	130,93	402,8	
01/02/02	6,55	62,13	848,55		6,55	40,83	523,36	
04/02/02	366,76	361,71	-1,38	X	366,76	431,54	17,66	X
05/02/02	119,28	153,95	29,07		119,28	114,28	-4,19	X
06/02/02	157,32	204,35	29,89		157,32	247,67	57,43	
07/02/02	187,82	109,24	-41,84		187,82	77,69	-58,64	
08/02/02	161,37	112,57	-30,24		161,37	24,69	-84,7	
13/02/02	1208,38	343,09	-71,61		1208,38	411,1	-65,98	
14/02/02	494,52	381,97	-22,76		494,52	160,08	-67,63	
15/02/02	301,38	22,18	-92,64		301,38	271,09	-10,05	X
18/02/02	764,54	193,31	-74,72		764,54	184,16	-75,91	
19/02/02	212,45	301,87	42,09		212,45	167,3	-21,25	
20/02/02	202,51	620,98	206,64		202,51	810,3	300,13	
21/02/02	146,53	194,9	33,01		146,53	306,13	108,92	
22/02/02	154,65	229,08	48,13		154,65	279,3	80,6	
25/02/02	294,26	394,53	34,08		294,26	395,17	34,29	
26/02/02	-40,6	166,56	-510,25		-40,6	139,55	-443,72	
27/02/02	8,41	306,55	3545,07		8,41	516,73	6044,23	
28/02/02	-128,12	142,16	-210,96		-128,12	196,62	-253,47	
01/03/02	-96,37	130,28	-235,19		-96,37	172,17	-278,66	
04/03/02	242,53	255,86	5,5	X	242,53	290,62	19,83	X
05/03/02	70,68	105,17	48,8		70,68	3,79	-94,64	
06/03/02	-27,16	176,45	-749,67		-27,16	267,75	-1085,82	
07/03/02	152,57	30,14	-80,25		152,57	25,07	-83,57	
08/03/02	34,25	122,02	256,26		34,25	47,19	37,78	
11/03/02	291,27	237,72	-18,39	X	291,27	263,4	-9,57	X
12/03/02	178,24	136,57	-23,38		178,24	26,01	-85,41	
13/03/02	126,19	121,84	-3,45	X	126,19	154,09	22,11	
14/03/02	108,76	146,85	35,02		108,76	88,34	-18,78	X
15/03/02	145,5	91,66	-37		145,5	50,34	-65,4	
18/03/02	426,93	275,09	-35,57		426,93	287,2	-32,73	
19/03/02	192,4	209,11	8,69	X	192,4	103,03	-46,45	
20/03/02	110,46	140,42	27,12		110,46	159,21	44,13	
21/03/02	101,27	142,17	40,39		101,27	111,06	9,67	X
22/03/02	-57,28	138,28	-341,41		-57,28	100,01	-274,6	
25/03/02	281,24	261,98	-6,85	X	281,24	341,49	21,42	
26/03/02	-90,7	195,58	-315,63		-90,7	128,15	-241,29	
27/03/02	-180,29	74,64	-141,4		-180,29	115,86	-164,26	
28/03/02	-32,21	80,89	-351,13		-32,21	70,49	-318,85	
			Total Ac.	20%			Total Ac.	22%

Quadro 15: Previsão da série da agência 13 (sem outliers) segundo os modelos de Box-Jenkins e amortecimento exponencial.

Data	Agência 13 - Box-Jenkins				Agência 13 - Amortecimento exponencial			
	Real	Previsto	Dif. %	Acertos	Real	Previsto	Dif. %	Acertos
16/01/02	26,61	34,98	31,45		26,61	5,81	-78,17	
17/01/02	6,08	24,51	303,13		6,08	-0,07	-101,15	
18/01/02	-8,98	-8,04	-10,47	X	-8,98	-7,53	-16,15	X
21/01/02	139,15	26,34	-81,07		139,15	-5,78	-104,15	
22/01/02	52,24	12,64	-75,8		52,24	-11,78	-122,55	
23/01/02	15,72	10,55	-32,89		15,72	8,19	-47,9	
24/01/02	62,09	38,34	-38,25		62,09	2,08	-96,65	
25/01/02	-66,13	10,36	-115,67		-66,13	-4,91	-92,58	
28/01/02	18,16	19,17	5,56	X	18,16	-2,86	-115,75	
29/01/02	3,47	11,99	245,53		3,47	-10,55	-404,03	
30/01/02	2,26	36,74	1525,66		2,26	8,62	281,42	
31/01/02	-25,77	-6,71	-73,96		-25,77	2,7	-110,48	
01/02/02	-49,89	-23,29	-53,32		-49,89	-5,88	-88,21	
04/02/02	-14,47	34,65	-339,46		-14,47	-3,15	-78,23	
05/02/02	-20,25	23,03	-213,73		-20,25	-11,21	-44,64	
06/02/02	-13,66	-0,31	-97,73		-13,66	7,64	-155,93	
07/02/02	-2,5	16,38	-755,2		-2,5	1,44	-157,6	
08/02/02	-12,54	-0,54	-95,69		-12,54	7	-155,82	
13/02/02	138,45	19,59	-85,85		138,45	-3,7	-102,67	
14/02/02	-4,04	26,44	-754,46		-4,04	-10,65	163,61	
15/02/02	16,43	5,22	-68,23		16,43	8,82	-46,32	
18/02/02	46,97	7,33	-84,39		46,97	2,96	-93,7	
19/02/02	18,07	9,88	-45,32		18,07	-5,74	-131,77	
20/02/02	12,37	59,88	384,07		12,37	-0,77	-106,22	
21/02/02	12,19	12,96	6,32	X	12,19	-9,67	-179,33	
22/02/02	-42,43	15,38	-136,25		-42,43	10	-123,57	
25/02/02	-35,15	18,79	-153,46		-35,15	3,76	-110,7	
26/02/02	-45,38	-8,86	-80,48		-45,38	-5,92	-86,95	
27/02/02	-17,5	9,87	-156,4		-17,5	-1,66	-90,51	
28/02/02	-15,11	3,51	-123,23		-15,11	-10,79	-28,59	
01/03/02	-71,32	-6,54	-90,83		-71,32	8,12	-111,39	
04/03/02	-57,6	-6,77	-88,25		-57,6	1,61	-102,8	
05/03/02	-0,6	-6,7	1016,67		-0,6	-7,51	1151,67	
06/03/02	-7,56	-1,27	-83,2		-7,56	-3,1	-58,99	
07/03/02	-25,72	-2,65	-89,7		-25,72	-11,8	-54,12	
08/03/02	-2,33	-11,57	396,57		-2,33	6,08	-360,94	
11/03/02	38,38	-4,93	-112,85		38,38	0,52	-98,65	
12/03/02	27,58	13,41	-51,38		27,58	-7,28	-126,4	
13/03/02	18,59	36,73	97,58		18,59	-2,68	-114,42	
14/03/02	14,09	2,29	-83,75		14,09	-11,2	-179,49	
15/03/02	24,64	8,88	-63,96		24,64	7,1	-71,19	
18/03/02	55,39	25,34	-54,25		55,39	2,06	-96,28	
19/03/02	71,28	24,08	-66,22		71,28	-6,14	-108,61	
20/03/02	51,44	6,2	-87,95		51,44	-0,36	-100,7	
21/03/02	-40,39	18,94	-146,89		-40,39	-8,44	-79,1	
22/03/02	-32,34	-4,12	-87,26		-32,34	8,91	-127,55	
25/03/02	20,8	5,29	-74,57		20,8	3,59	-82,74	
26/03/02	-9,43	19,63	-308,17		-9,43	-3,89	-58,75	
27/03/02	23,63	12,95	-45,2		23,63	-0,17	-100,72	
28/03/02	17,55	0,7	-96,01		17,55	-8,93	-150,88	
			Tot. Ac.	6%			Tot. Ac.	2%

4. MODELOS E EXPERIMENTOS REALIZADOS – EXPERIMENTO 1

Quadro 16: Quadro resumo das melhores topologias nos testes de redes neurais utilizando 4 formas de normalizações lineares.

Tipo de normalização	Agência 02		Agência 06		Agência 13	
	Topologia	% de acertos	Topologia	% de acertos	Topologia	% de acertos
1	50 x 5 x 1	17,1	20 x 10 x 1	12,0	10 x 10 x 1	11,0
					20 x 5 x 1	11,0
2	20 x 10 x 1	12,1	10 x 5 x 1	15,0	20 x 10 x 1	13,0
3	10 x 5 x 1	15,0	10 x 10 x 1	13,5	10 x 5 x 1	13,6
			20 x 10 x 1	13,5		
4	10 x 10 x 1	11,4	10 x 5 x 1	9,8	30 x 10 x 1	13,0
			10 x 10 x 1	9,8		
			20 x 10 x 1	9,8		
			40 x 10 x 1	9,8		
			50 x 10 x 1	9,8		

Quadro 17: Resultados relativos aos treinamentos com as melhores redes.

Agência 02		Agência 06		Agência 13	
Quant.	% Acertos	Quant.	% Acertos	Quant.	% Acertos
17	11,4	12	8,1	16	10,7
16	10,7	12	8,1	17	11,4
8	5,4	14	9,4	14	9,4
12	8,1	19	12,8	10	6,7
8	5,4	9	6,0	16	10,7

Quadro 18: Acertos na direção dos dados previstos pela melhor rede para d e $d - 1$.

	Acertos (Direção)	% Acertos	Erros (Direção)	% Erros
Ag. 02	56	37,84	92	62,16
Ag. 06	70	47,30	78	52,70
Ag. 13	66	44,00	84	56,00

5. MODELOS E EXPERIMENTOS REALIZADOS – EXPERIMENTO 2

Quadro 19: Resultados do treinamento de redes usando a técnica de CPD para a série de dados da agência 02. Em destaque tem-se o melhor resultado alcançado.

Neuro	Quant. Acertos	% Acertos	Neuro	Quant. Acertos	% Acertos
1	22	14.8	26	6	4.0
2	15	10.1	27	8	5.4
3	11	7.4	28	6	4.0
4	13	8.7	29	4	2.7
5	13	8.7	30	7	4.7
6	12	8.1	31	9	6.0
7	11	7.4	32	7	4.7
8	12	8.1	33	9	6.0
9	11	7.4	34	8	5.4
10	11	7.4	35	6	4.0
11	11	7.4	36	8	5.4
12	5	3.4	37	5	3.4
13	7	4.7	38	9	6.0
14	9	6.0	39	6	4.0
15	10	6.7	40	8	5.4
16	9	6.0	41	4	2.7
17	8	5.4	42	8	5.4
18	10	6.7	43	8	5.4
19	7	4.7	44	10	6.7
20	7	4.7	45	7	4.7
21	7	4.7	46	7	4.7
22	7	4.7	47	9	6.0
23	7	4.7	48	7	4.7
24	8	5.4	49	7	4.7
25	8	5.4	50	9	6.0

Quadro 20: Resultados do treinamento de redes usando a técnica de CPD para a série de dados da agência 06. Em destaque têm-se os melhores resultados alcançados.

Neuro	Quant. Acertos	% Acertos	Neuro	Quant. Acertos	% Acertos
1	5	3.4	26	10	6.7
2	8	5.4	27	11	7.4
3	8	5.4	28	12	8.1
4	5	3.4	29	12	8.1
5	8	5.4	30	12	8.1
6	8	5.4	31	13	8.7
7	8	5.4	32	14	9.4
8	9	6.0	33	13	8.7
9	8	5.4	34	15	10.1
10	10	6.7	35	15	10.1
11	11	7.4	36	18	12.1
12	10	6.7	37	12	8.1
13	9	6.0	38	14	9.4
14	8	5.4	39	16	10.7
15	9	6.0	40	17	11.4
16	5	3.4	41	13	8.7
17	5	3.4	42	18	12.1
18	8	5.4	43	16	10.7
19	8	5.4	44	14	9.4
20	11	7.4	45	15	10.1
21	8	5.4	46	9	6.0
22	8	5.4	47	16	10.7
23	9	6.0	48	11	7.4
24	7	4.7	49	18	12.1
25	10	6.7	50	18	12.1

Quadro 21: Resultados do treinamento de redes usando a técnica de CPD para a série de dados da agência 13. Em destaque têm-se os melhores resultados alcançados.

Neuro	Quant. Acertos	% Acertos	Neuro	Quant. Acertos	% Acertos
1	16	10.7	26	12	8.1
2	16	10.7	27	14	9.4
3	17	11.4	28	13	8.7
4	18	12.1	29	14	9.4
5	20	13.4	30	12	8.1
6	19	12.8	31	12	8.1
7	17	11.4	32	13	8.7
8	16	10.7	33	13	8.7
9	13	8.7	34	13	8.7
10	12	8.1	35	13	8.7
11	11	7.4	36	12	8.1
12	16	10.7	37	14	9.4
13	15	10.1	38	13	8.7
14	11	7.4	39	11	7.4
15	15	10.1	40	11	7.4
16	13	8.7	41	14	9.4
17	13	8.7	42	14	9.4
18	13	8.7	43	9	6.0
19	14	9.4	44	14	9.4
20	13	8.7	45	11	7.4
21	14	9.4	46	17	11.4
22	15	10.1	47	15	10.1
23	14	9.4	48	10	6.7
24	13	8.7	49	12	8.1
25	12	8.1	50	13	8.7

6. MODELOS E EXPERIMENTOS REALIZADOS – EXPERIMENTO 3

Quadro 22: Resultados do treinamento de redes usando a técnica de CPD para a série de dados da agência 02. Em destaque tem-se o melhor resultado alcançado.

Neuro	Quant. Acertos	% Acertos	Neuro	Quant. Acertos	% Acertos
1	9	6.3	26	13	9.0
2	13	9.0	27	15	10.4
3	10	6.9	28	15	10.4
4	10	6.9	29	18	12.5
5	17	11.8	30	14	9.7
6	14	9.7	31	19	13.2
7	20	13.9	32	20	13.9
8	18	12.5	33	18	12.5
9	18	12.5	34	19	13.2
10	16	11.1	35	19	13.2
11	14	9.7	36	20	13.9
12	14	9.7	37	21	14.6
13	11	7.6	38	23	16.0
14	15	10.4	39	15	10.4
15	13	9.0	40	19	13.2
16	16	11.1	41	16	11.1
17	15	10.4	42	15	10.4
18	19	13.2	43	19	13.2
19	14	9.7	44	18	12.5
20	9	6.3	45	18	12.5
21	10	6.9	46	10	6.9
22	13	9.0	47	16	11.1
23	13	9.0	48	17	11.8
24	12	8.3	49	14	9.7
25	13	9.0	50	17	11.8

Quadro 23: Resultados do treinamento de redes usando a técnica de CPD para a série de dados da agência 06. Em destaque têm-se os melhores resultados alcançados.

Neuro	Acertos (20%)	% Acertos	Neuro	Acertos (20%)	% Acertos
1	10	6.9	26	9	6.3
2	18	12.5	27	8	5.6
3	17	11.8	28	14	9.7
4	11	7.6	29	14	9.7
5	13	9.0	30	12	8.3
6	10	6.9	31	16	11.1
7	12	8.3	32	14	9.7
8	10	6.9	33	15	10.4
9	10	6.9	34	16	11.1
10	13	9.0	35	13	9.0
11	11	7.6	36	16	11.1
12	10	6.9	37	12	8.3
13	12	8.3	38	15	10.4
14	12	8.3	39	14	9.7
15	14	9.7	40	12	8.3
16	13	9.0	41	14	9.7
17	12	8.3	42	9	6.3
18	12	8.3	43	11	7.6
19	13	9.0	44	11	7.6
20	12	8.3	45	11	7.6
21	11	7.6	46	12	8.3
22	13	9.0	47	9	6.3
23	12	8.3	48	14	9.7
24	13	9.0	49	17	11.8
25	11	7.6	50	12	8.3

Quadro 24: Resultados do treinamento de redes usando a técnica de CPD para a série de dados da agência 13. Em destaque tem-se o melhor resultado alcançado.

Neuro	Quant. Acertos	% Acertos	Neuro	Quant. Acertos	% Acertos
1	15	10.4	26	14	9.7
2	20	13.9	27	11	7.6
3	12	8.3	28	10	6.9
4	13	9.0	29	21	14.6
5	13	9.0	30	14	9.7
6	13	9.0	31	16	11.1
7	14	9.7	32	11	7.6
8	14	9.7	33	13	9.0
9	15	10.4	34	14	9.7
10	11	7.6	35	19	13.2
11	12	8.3	36	17	11.8
12	13	9.0	37	16	11.1
13	13	9.0	38	12	8.3
14	10	6.9	39	14	9.7
15	11	7.6	40	14	9.7
16	11	7.6	41	14	9.7
17	11	7.6	42	16	11.1
18	11	7.6	43	15	10.4
19	10	6.9	44	13	9.0
20	10	6.9	45	18	12.5
21	11	7.6	46	13	9.0
22	8	5.6	47	14	9.7
23	9	6.3	48	16	11.1
24	10	6.9	49	17	11.8
25	11	7.6	50	19	13.2

7. MODELOS E EXPERIMENTOS REALIZADOS – EXPERIMENTO 5

Quadro 25: Agência 02 - Resultados do treinamento de redes Elman. Em destaque tem-se o melhor resultado alcançado.

Neuro	Quant. acertos	% Acertos	Neuro	Quant. acertos	% Acertos
1	6	3.9	26	16	10.5
2	10	6.6	27	17	11.2
3	13	8.6	28	13	8.6
4	18	11.8	29	13	8.6
5	11	7.2	30	13	8.6
6	25	16.4	31	18	11.8
7	20	13.2	32	12	7.9
8	19	12.5	33	13	8.6
9	9	5.9	34	15	9.9
10	13	8.6	35	16	10.5
11	12	7.9	36	17	11.2
12	16	10.5	37	18	11.8
13	18	11.8	38	19	12.5
14	20	13.2	39	10	6.6
15	15	9.9	40	14	9.2
16	14	9.2	41	16	10.5
17	19	12.5	42	12	7.9
18	18	11.8	43	16	10.5
19	18	11.8	44	12	7.9
20	19	12.5	45	16	10.5
21	11	7.2	46	15	9.9
22	20	13.2	47	13	8.6
23	15	9.9	48	19	12.5
24	14	9.2	49	14	9.2
25	22	14.5	50	18	11.8

Quadro 26: Agência 06 - Resultados do treinamento de redes Elman. Em destaque tem-se o melhor resultado alcançado.

Neuro	Quant. acertos	% Acertos	Neuro	Quant. acertos	% Acertos
1	14	9.2	26	13	8.6
2	10	6.6	27	14	9.2
3	14	9.2	28	15	9.9
4	16	10.5	29	19	12.5
5	13	8.6	30	19	12.5
6	9	5.9	31	16	10.5
7	13	8.6	32	14	9.2
8	13	8.6	33	15	9.9
9	14	9.2	34	13	8.6
10	15	9.9	35	15	9.9
11	16	10.5	36	16	10.5
12	17	11.2	37	16	10.5
13	21	13.8	38	17	11.2
14	17	11.2	39	17	11.2
15	21	13.8	40	19	12.5
16	22	14.5	41	13	8.6
17	15	9.9	42	14	9.2
18	14	9.2	43	18	11.8
19	16	10.5	44	14	9.2
20	17	11.2	45	16	10.5
21	16	10.5	46	16	10.5
22	17	11.2	47	16	10.5
23	17	11.2	48	17	11.2
24	16	10.5	49	11	7.2
25	16	10.5	50	21	13.8

Quadro 27: Agência 13 - Resultados do treinamento de redes Elman. Em destaque tem-se o melhor resultado alcançado.

Neuro	Quant. acertos	% Acertos	Neuro	Quant. acertos	% Acertos
1	13	8,6	26	12	7,9
2	15	9,9	27	16	10,5
3	17	11,2	28	12	7,9
4	13	8,6	29	10	6,6
5	13	8,6	30	14	9,2
6	12	7,9	31	11	7,2
7	15	9,9	32	18	11,8
8	14	9,2	33	11	7,2
9	15	9,9	34	7	4,6
10	8	5,3	35	15	9,9
11	10	6,6	36	14	9,2
12	10	6,6	37	13	8,6
13	10	6,6	38	17	11,2
14	11	7,2	39	16	10,5
15	16	10,5	40	16	10,5
16	15	9,9	41	15	9,9
17	17	11,2	42	16	10,5
18	12	7,9	43	12	7,9
19	13	8,6	44	15	9,9
20	14	9,2	45	17	11,2
21	8	5,3	46	11	7,2
22	10	6,6	47	15	9,9
23	12	7,9	48	14	9,2
24	9	5,9	49	14	9,2
25	12	7,9	50	14	9,2

8. MODELOS E EXPERIMENTOS REALIZADOS – EXPERIMENTO 6

Quadro 28: Resumo dos melhores resultados no treinamento de redes MLP e Elman para as agências 02, 06 e 13, com 1 a 50 neurônios na primeira camada e utilizando 4 formas de normalizações não lineares.

Norma	Agência 02				Agência 06				Agência 13			
	MLP		Elman		MLP		Elman		MLP		Elman	
	Neuro	% acertos	Neuro	% acertos	Neuro	% acertos	Neuro	% acertos	Neuro	% acertos	Neuro	% acertos
1	6	19,1	38	17,1	13	23,7	49	17,8	4	21,1	4	19,1
2	4	9,2	19	5,9	14	13,2	16	11,2	5	15,1	3	11,2
3	11	23,0	13	25,7	10	29,6	12	30,3	32	7,9	1	14,5
4	2	19,7	12	23,0	3	26,3	12	30,3	13	19,1	50	17,1

Quadro 29: Agência 02 - Treinamento das melhores redes de MLP e Elman usadas na primeira fase.

		Agência 02					
		Rede MLP			Rede Elman		
		Neuro	Quant. acertos	% Acertos	Neuro	Quant. acertos	% Acertos
Tipo de normalização	1	4	26	17,1	26	22	14,5
		5	22	14,5	27	22	14,5
		6	22	14,5	28	22	14,5
		7	22	14,5	29	0	0,0
		8	22	14,5	30	22	14,5
	2	1	18	11,8	19	2	1,3
		2	5	3,3	20	5	3,3
		3	4	2,6	21	5	3,3
		4	3	2,0	22	6	3,9
		5	4	2,6	23	5	3,3
	3	10	16	10,5	11	28	18,4
		11	22	14,5	12	28	18,4
		12	24	15,8	13	29	19,1
		13	23	15,1	14	25	16,4
		14	21	13,8	15	30	19,7
	4	1	27	17,8	10	24	15,8
		2	21	13,8	11	28	18,4
		3	35	23,0	12	31	20,4
		4	30	19,7	13	26	17,1
		5	24	15,8	14	29	19,1

Quadro 30: Agência 06 - Treinamento das melhores redes de MLP e Elman usadas na primeira fase.

		Agência 06					
		Rede MLP			Rede Elman		
		Neuro	Quant. acertos	% Acertos	Neuro	Quant. acertos	% Acertos
Tipo de normalização	1	12	36	23,7	46	17	11,2
		13	37	24,3	47	32	21,1
		14	17	11,2	48	20	13,2
		15	37	24,3	49	18	11,8
		16	35	23,0	50	20	13,2
	2	13	0	0,0	15	5	3,3
		14	1	0,7	16	7	4,6
		15	0	0,0	17	12	7,9
		16	2	1,3	18	7	4,6
		17	0	0,0	19	5	3,3
	3	8	17	11,2	11	13	8,6
		9	16	10,5	12	19	12,5
		10	15	9,9	13	17	11,2
		11	22	14,5	14	16	10,5
		12	17	11,2	15	20	13,2
	4	2	17	11,2	36	21	13,8
		3	18	11,8	37	14	9,2
		4	11	7,2	38	13	8,6
		5	18	11,8	39	14	9,2
		6	17	11,2	40	14	9,2

Quadro 31: Agência 13 - Treinamento das melhores redes de MLP e Elman usadas na primeira fase.

		Agência 13					
		Rede MLP			Rede Elman		
		Neuro	Quant. acertos	% Acertos	Neuro	Quant. acertos	% Acertos
Tipo de normalização	1	1	25	16.4	1	25	16.4
		2	25	16.4	2	25	16.4
		3	25	16.4	3	25	16.4
		4	32	21.1	4	25	16.4
		5	31	20.4	5	25	16.4
	2	4	1	0.7	2	4	2.6
		5	0	0.0	3	3	2.0
		6	0	0.0	4	2	1.3
		7	0	0.0	5	0	0.0
		8	0	0.0	6	2	1.3
	3	30	2	1.3	8	5	3.3
		31	3	2.0	9	4	2.6
		32	1	0.7	10	3	2.0
		33	1	0.7	11	8	5.3
		34	2	1.3	12	3	2.0
	4	12	0	0.0	1	5	3.3
13		5	3.3	2	7	4.6	
14		3	2.0	3	7	4.6	
15		4	2.6	4	7	4.6	
16		3	2.0	5	7	4.6	

9. MODELOS E EXPERIMENTOS REALIZADOS – EXPERIMENTO 7

Quadro 32: Resumo das melhores redes MLP e Elman treinadas para as agências 02 usando funções de propagação não lineares nas duas camadas.

Agência 02						
Tipo de normalização	Rede MLP			Rede Elman		
	Arquitetura	Quant. acertos	% acertos	Arquitetura	Quant. acertos	% acertos
1	-	-	-	-	-	-
2	12 x 1	24	15,8	3 x 1	20	13,2
3	10 x 1	23	15,1	3 x 1 e 38 x 1	30	19,7
4	7 x 1	37	24,3	16 x 1	37	24,3

Quadro 33: Resumo das melhores redes MLP e Elman treinadas para as agências 06 usando funções de propagação não lineares nas duas camadas.

Agência 06						
Tipo de normalização	Rede MLP			Rede Elman		
	Arquitetura	Quant. acertos	% acertos	Arquitetura	Quant. acertos	% acertos
1	-	-	-	-	-	-
2	6 x 1	20	13,2	25 x 1	27	17,8
3	4 x 1	23	15,1	21 x 1 e 32 x 1	23	15,1
4	9 x 1 e 26 x 1 e 38 x 1	23	15,1	10 x 1 e 13 x 1	23	15,1

Quadro 34: Resumo das melhores redes MLP e Elman treinadas para as agências 13 usando funções de propagação não lineares nas duas camadas.

Agência 13						
Tipo de normalização	Rede MLP			Rede Elman		
	Arquitetura	Quant. acertos	% acertos	Arquitetura	Quant. acertos	% acertos
1	-	-	-	-	-	-
2	10 x 1	25	16,5	6 x 1	24	15,8
3	23 x 1	12	7,9	13 x 1	24	15,8
4	9 x 1	27	17,8	10 x 1	28	18,4

10. MODELOS E EXPERIMENTOS REALIZADOS – EXPERIMENTO 8

Quadro 35: Agência 02 – Performance da rede em 4 e 5 intervalos.

Intervalo 4			Intervalo 5		
Neuro	Qtd. Acertos	% Acertos	Neuro	Qtd. Acertos	% Acertos
1	114	72,61	1	69	43,95
2	114	72,61	2	52	33,12
3	114	72,61	3	52	33,12
4	114	72,61	4	52	33,12
5	114	72,61	5	52	33,12
6	114	72,61	6	69	43,95
7	114	72,61	7	73	46,5
8	114	72,61	8	70	44,59
9	109	69,43	9	70	44,59
10	110	70,06	10	64	40,76
11	109	69,43	11	69	43,95
12	112	71,34	12	66	42,04
13	113	71,97	13	73	46,5
14	109	69,43	14	65	41,4
15	106	67,52	15	71	45,22
16	113	71,97	16	72	45,86
17	111	70,7	17	73	46,5
18	113	71,97	18	75	47,77
19	103	65,61	19	69	43,95
20	106	67,52	20	70	44,59

Quadro 36: Agência 02 – Performance da rede em 6 e 8 intervalos.

Intervalo 6			Intervalo 8		
Neuro	Qtd. Acertos	% Acertos	Neuro	Qtd. Acertos	% Acertos
1	78	49,68	1	58	36,94
2	94	59,87	2	58	36,94
3	94	59,87	3	58	36,94
4	94	59,87	4	58	36,94
5	94	59,87	5	58	36,94
6	78	49,68	6	72	45,86
7	94	59,87	7	69	43,95
8	86	54,78	8	58	36,94
9	87	55,41	9	76	48,41
10	89	56,69	10	73	46,5
11	89	56,69	11	62	39,49
12	89	56,69	12	69	43,95
13	86	54,78	13	70	44,59
14	87	55,41	14	75	47,77
15	95	60,51	15	70	44,59
16	87	55,41	16	68	43,31
17	87	55,41	17	70	44,59
18	76	48,41	18	70	44,59
19	79	50,32	19	70	44,59
20	84	53,5	20	71	45,22

Quadro 37: Ag. 06 – Performance da rede em 4 e 5 intervalos.

Intervalo 4			Intervalo 5		
Neuro	Qtd. Acertos	% Acertos	Neuro	Qtd. Acertos	% Acertos
1	72	45,86	1	99	63,06
2	72	45,86	2	99	63,06
3	72	45,86	3	99	63,06
4	72	45,86	4	99	63,06
5	72	45,86	5	99	63,06
6	72	45,86	6	99	63,06
7	72	45,86	7	99	63,06
8	72	45,86	8	99	63,06
9	73	46,5	9	99	63,06
10	89	56,69	10	99	63,06
11	79	50,32	11	91	57,96
12	86	54,78	12	83	52,87
13	81	51,59	13	84	53,5
14	71	45,22	14	80	50,96
15	89	56,69	15	86	54,78
16	87	55,41	16	85	54,14
17	81	51,59	17	86	54,78
18	87	55,41	18	78	49,68
19	88	56,05	19	88	56,05
20	88	56,05	20	80	50,96

Quadro 38: Agência 06 – Performance da rede em 6 e 8 intervalos.

Intervalo 6			Intervalo 8		
Neuro	Qtd. Acertos	% Acertos	Neuro	Qtd. Acertos	% Acertos
1	98	62,42	1	76	48,41
2	98	62,42	2	76	48,41
3	98	62,42	3	76	48,41
4	98	62,42	4	76	48,41
5	97	61,78	5	76	48,41
6	95	60,51	6	75	47,77
7	98	62,42	7	76	48,41
8	98	62,42	8	76	48,41
9	52	33,12	9	79	50,32
10	98	62,42	10	76	48,41
11	92	58,6	11	77	49,04
12	88	56,05	12	78	49,68
13	93	59,24	13	72	45,86
14	87	55,41	14	75	47,77
15	85	54,14	15	76	48,41
16	84	53,5	16	72	45,86
17	84	53,5	17	72	45,86
18	90	57,32	18	75	47,77
19	91	57,96	19	77	49,04
20	87	55,41	20	73	46,5

Quadro 39: Ag. 13 – Performance da rede em 4 e 5 intervalos.

Intervalo 4			Intervalo 5		
Neuro	Qtd. Acertos	% Acertos	Neuro	Qtd. Acertos	% Acertos
1	88	56,05	1	118	75,16
2	88	56,05	2	118	75,16
3	88	56,05	3	118	75,16
4	88	56,05	4	118	75,16
5	88	56,05	5	118	75,16
6	85	54,14	6	118	75,16
7	77	49,04	7	118	75,16
8	77	49,04	8	118	75,16
9	82	52,23	9	112	71,34
10	81	51,59	10	115	73,25
11	75	47,77	11	112	71,34
12	83	52,87	12	112	71,34
13	80	50,96	13	113	71,97
14	82	52,23	14	103	65,61
15	80	50,96	15	109	69,43
16	76	48,41	16	115	73,25
17	79	50,32	17	106	67,52
18	83	52,87	18	107	68,15
19	81	51,59	19	113	71,97
20	92	58,6	20	100	63,69

Quadro 40: Agência 13 – Performance da rede em 6 e 8 intervalos.

Intervalo 6			Intervalo 8		
Neuro	Qtd. Acertos	% Acertos	Neuro	Qtd. Acertos	% Acertos
1	73	46,5	1	70	44,59
2	73	46,5	2	58	36,94
3	73	46,5	3	70	44,59
4	73	46,5	4	70	44,59
5	73	46,5	5	61	38,85
6	72	45,86	6	62	39,49
7	85	54,14	7	59	37,58
8	85	54,14	8	56	35,67
9	82	52,23	9	59	37,58
10	74	47,13	10	58	36,94
11	82	52,23	11	66	42,04
12	82	52,23	12	61	38,85
13	80	50,96	13	71	45,22
14	78	49,68	14	67	42,68
15	79	50,32	15	56	35,67
16	75	47,77	16	60	38,22
17	69	43,95	17	60	38,22
18	78	49,68	18	54	34,39
19	74	47,13	19	58	36,94
20	72	45,86	20	57	36,31

11. MODELOS E EXPERIMENTOS REALIZADOS – EXPERIMENTO 9

Quadro 41: Performance alcançada por intervalo.

Intervalos	Ag. 02	Ag. 06	Ag. 13
	rede 3 x 1	rede 7 x 1	rede 10 x 1
2	94,27	61,15	84,08
3	63,69	78,34	71,97
4	71,34	31,21	67,52
5	52,87	70,70	67,52
6	57,96	35,67	49,04
7	37,58	47,13	45,86
8	41,40	22,93	47,77
9	37,58	42,68	30,57
10	36,31	27,39	33,12

12. MODELOS E EXPERIMENTOS REALIZADOS – EXPERIMENTO 10

Quadro 42:Ag. 02 – Performance binarização dos dados.

Agência 02				
Qtd neurônios	MLP		Elman	
	Qtd. Acertos	% Acertos	Qtd. Acertos	% Acertos
1	117	74.52	117	74.52
2	117	74.52	111	70.70
3	107	68.15	104	66.24
4	114	72.61	117	74.52
5	110	70.06	105	66.88
6	108	68.79	110	70.06
7	111	70.70	110	70.06
8	102	64.97	105	66.88
9	107	68.15	106	67.52
10	107	68.15	102	64.97
11	103	65.61	110	70.06
12	104	66.24	110	70.06
13	110	70.06	100	63.69
14	108	68.79	101	64.33
15	102	64.97	104	66.24
16	106	67.52	107	68.15
17	102	64.97	102	64.97
18	102	64.97	101	64.33
19	105	66.88	102	64.97
20	104	66.24	99	63.06

Quadro 43:Ag. 06 – Performance binarização dos dados.

Agência 06				
Qtd neurônios	MLP		Elman	
	Qtd. Acertos	% Acertos	Qtd. Acertos	% Acertos
1	119	75.80	119	75.80
2	117	74.52	115	73.25
3	110	70.06	114	72.61
4	108	68.79	108	68.79
5	119	75.80	109	69.43
6	104	66.24	106	67.52
7	109	69.43	107	68.15
8	108	68.79	109	69.43
9	109	69.43	113	71.97
10	111	70.70	107	68.15
11	111	70.70	107	68.15
12	107	68.15	109	69.43
13	109	69.43	106	67.52
14	107	68.15	110	70.06
15	108	68.79	109	69.43
16	110	70.06	113	71.97
17	108	68.79	109	69.43
18	108	68.79	111	70.70
19	110	70.06	108	68.79
20	103	65.61	112	71.34

Quadro 44:Ag. 13 – Performance binarização dos dados.

Agência 13				
Qtd neurônios	MLP		Elman	
	Qtd. Acertos	% Acertos	Qtd. Acertos	% Acertos
1	68	43.31	69	43.95
2	69	43.95	67	42.68
3	68	43.31	63	40.13
4	66	42.04	70	44.59
5	66	42.04	60	38.22
6	69	43.95	59	37.58
7	64	40.76	59	37.58
8	64	40.76	59	37.58
9	62	39.49	56	35.67
10	60	38.22	64	40.76
11	60	38.22	62	39.49
12	64	40.76	63	40.13
13	66	42.04	61	38.85
14	65	41.40	60	38.22
15	63	40.13	60	38.22
16	56	35.67	61	38.85
17	58	36.94	58	36.94
18	54	34.39	58	36.94
19	59	37.58	57	36.31
20	59	37.58	64	40.76

13. MODELOS E EXPERIMENTOS REALIZADOS – EXPERIMENTO 11

Quadro 45:Ag. 02 – Performance binarização dos dados – experimento 11.

Agência 02											
		Norma 1		Norma 2		Norma 3		Norma 4		Norma 5	
Neuro 1	Neuro 2	Qtd. Ac.	Perc. %								
1	4	89	56,69	90	57,32	89	56,69	87	55,41	64	40,76
2	4	90	57,32	89	56,69	88	56,05	83	52,87	66	42,04
3	4	88	56,05	94	59,87	83	52,87	56	35,67	64	40,76
4	4	68	43,31	91	57,96	89	56,69	83	52,87	69	43,95
5	4	80	50,96	89	56,69	96	61,15	88	56,05	66	42,04
6	4	87	55,41	87	55,41	90	57,32	94	59,87	63	40,13
7	4	77	49,04	74	47,13	93	59,24	87	55,41	68	43,31
8	4	83	52,87	75	47,77	92	58,6	97	61,78	62	39,49
9	4	74	47,13	83	52,87	82	52,23	90	57,32	68	43,31
10	4	85	54,14	81	51,59	67	42,68	96	61,15	68	43,31
11	4	74	47,13	64	40,76	87	55,41	97	61,78	65	41,4
12	4	81	51,59	83	52,87	79	50,32	94	59,87	62	39,49
13	4	74	47,13	71	45,22	82	52,23	93	59,24	53	33,76
14	4	78	49,68	77	49,04	86	54,78	91	57,96	68	43,31
15	4	79	50,32	89	56,69	68	43,31	91	57,96	61	38,85
16	4	77	49,04	67	42,68	84	53,5	92	58,6	58	36,94
17	4	79	50,32	66	42,04	74	47,13	92	58,6	60	38,22
18	4	74	47,13	75	47,77	82	52,23	99	63,06	70	44,59
19	4	82	52,23	71	45,22	74	47,13	94	59,87	78	49,68
20	4	69	43,95	81	51,59	77	49,04	98	62,42	60	38,22
21	4	74	47,13	62	39,49	78	49,68	90	57,32	56	35,67
22	4	82	52,23	79	50,32	77	49,04	96	61,15	52	33,12
23	4	67	42,68	66	42,04	74	47,13	94	59,87	63	40,13
24	4	79	50,32	75	47,77	82	52,23	89	56,69	64	40,76
25	4	77	49,04	72	45,86	83	52,87	93	59,24	48	30,57
26	4	82	52,23	72	45,86	82	52,23	93	59,24	54	34,39
27	4	74	47,13	65	41,4	84	53,5	94	59,87	55	35,03
28	4	73	46,5	71	45,22	77	49,04	101	64,33	72	45,86
29	4	77	49,04	64	40,76	89	56,69	96	61,15	76	48,41
30	4	66	42,04	67	42,68	68	43,31	92	58,6	63	40,13

Quadro 46:Ag. 06 – Performance binarização dos dados – experimento 11.

Agência 06											
		Norma 1		Norma 2		Norma 3		Norma 4		Norma 5	
Neuro 1	Neuro 2	Qtd. Ac.	Perc. %								
1	4	89	56,69	88	56,05	87	55,41	89	56,69	62	39,49
2	4	90	57,32	87	55,41	88	56,05	84	53,5	82	52,23
3	4	87	55,41	87	55,41	87	55,41	87	55,41	82	52,23
4	4	81	51,59	78	49,68	87	55,41	88	56,05	84	53,5
5	4	93	59,24	93	59,24	86	54,78	89	56,69	93	59,24
6	4	86	54,78	84	53,5	86	54,78	83	52,87	76	48,41
7	4	85	54,14	84	53,5	87	55,41	87	55,41	86	54,78
8	4	83	52,87	75	47,77	85	54,14	90	57,32	53	33,76
9	4	83	52,87	75	47,77	91	57,96	85	54,14	75	47,77
10	4	85	54,14	76	48,41	84	53,5	88	56,05	96	61,15
11	4	89	56,69	75	47,77	91	57,96	91	57,96	82	52,23
12	4	82	52,23	77	49,04	84	53,5	83	52,87	89	56,69
13	4	86	54,78	85	54,14	80	50,96	85	54,14	61	38,85
14	4	89	56,69	84	53,5	86	54,78	87	55,41	84	53,5
15	4	77	49,04	81	51,59	81	51,59	88	56,05	59	37,58
16	4	79	50,32	79	50,32	86	54,78	91	57,96	76	48,41
17	4	73	46,5	88	56,05	84	53,5	86	54,78	67	42,68
18	4	85	54,14	85	54,14	88	56,05	86	54,78	92	58,6
19	4	87	55,41	81	51,59	85	54,14	91	57,96	80	50,96
20	4	93	59,24	78	49,68	82	52,23	84	53,5	82	52,23
21	4	82	52,23	79	50,32	84	53,5	91	57,96	79	50,32
22	4	74	47,13	87	55,41	82	52,23	86	54,78	91	57,96
23	4	78	49,68	74	47,13	83	52,87	85	54,14	69	43,95
24	4	80	50,96	85	54,14	85	54,14	86	54,78	87	55,41
25	4	76	48,41	75	47,77	83	52,87	90	57,32	86	54,78
26	4	85	54,14	82	52,23	85	54,14	84	53,5	85	54,14
27	4	72	45,86	78	49,68	88	56,05	87	55,41	74	47,13
28	4	75	47,77	76	48,41	83	52,87	84	53,5	81	51,59
29	4	82	52,23	76	48,41	81	51,59	85	54,14	78	49,68
30	4	87	55,41	77	49,04	82	52,23	90	57,32	73	46,5

Quadro 47:Ag. 13 – Performance binarização dos dados – experimento 11.

Agência 13											
		Norma 1		Norma 2		Norma 3		Norma 4		Norma 5	
Neuro 1	Neuro 2	Qtd. Ac.	Perc. %								
1	4	68	43,31	76	48,41	75	47,77	73	46,5	62	39,49
2	4	70	44,59	70	44,59	76	48,41	62	39,49	62	39,49
3	4	70	44,59	72	45,86	73	46,5	74	47,13	60	38,22
4	4	72	45,86	71	45,22	76	48,41	71	45,22	65	41,4
5	4	70	44,59	70	44,59	64	40,76	77	49,04	63	40,13
6	4	70	44,59	77	49,04	77	49,04	72	45,86	59	37,58
7	4	62	39,49	76	48,41	74	47,13	81	51,59	63	40,13
8	4	67	42,68	70	44,59	64	40,76	69	43,95	56	35,67
9	4	57	36,31	70	44,59	63	40,13	67	42,68	61	38,85
10	4	61	38,85	73	46,5	60	38,22	65	41,4	59	37,58
11	4	69	43,95	69	43,95	62	39,49	70	44,59	56	35,67
12	4	65	41,4	61	38,85	65	41,4	69	43,95	62	39,49
13	4	65	41,4	62	39,49	67	42,68	68	43,31	61	38,85
14	4	71	45,22	65	41,4	68	43,31	68	43,31	56	35,67
15	4	59	37,58	59	37,58	72	45,86	70	44,59	55	35,03
16	4	61	38,85	71	45,22	65	41,4	70	44,59	54	34,39
17	4	68	43,31	56	35,67	70	44,59	69	43,95	49	31,21
18	4	66	42,04	67	42,68	69	43,95	67	42,68	58	36,94
19	4	60	38,22	57	36,31	68	43,31	70	44,59	50	31,85
20	4	64	40,76	64	40,76	64	40,76	65	41,4	64	40,76
21	4	65	41,4	67	42,68	67	42,68	66	42,04	56	35,67
22	4	61	38,85	66	42,04	67	42,68	66	42,04	65	41,4
23	4	62	39,49	66	42,04	67	42,68	65	41,4	56	35,67
24	4	61	38,85	58	36,94	72	45,86	67	42,68	54	34,39
25	4	61	38,85	76	48,41	71	45,22	63	40,13	57	36,31
26	4	62	39,49	68	43,31	63	40,13	68	43,31	63	40,13
27	4	63	40,13	71	45,22	62	39,49	63	40,13	56	35,67
28	4	69	43,95	61	38,85	70	44,59	65	41,4	65	41,4
29	4	60	38,22	60	38,22	63	40,13	64	40,76	52	33,12
30	4	61	38,85	71	45,22	66	42,04	61	38,85	52	33,12

14. MODELOS E EXPERIMENTOS REALIZADOS – EXPERIMENTO 12

Quadro 48: Agência 02 – Performance utilizando média móvel – experimento 12.

Agência 02									
MM 05		MM 10		MM 15		MM 20		Sem MM	
Acertos	Perc. %	Acertos	Perc. %	Acertos	Perc. %	Acertos	Perc. %	Acertos	Perc. %
2	0,65	19	6,19	21	6,84	0	0	22	7,17
7	2,28	3	0,98	2	0,65	0	0	27	8,79
15	4,89	1	0,33	2	0,65	1	0,33	13	4,23
7	2,28	3	0,98	0	0	0	0	19	6,19
5	1,63	4	1,3	0	0	0	0	20	6,51
5	1,63	4	1,3	5	1,63	1	0,33	13	4,23
3	0,98	14	4,56	1	0,33	0	0	11	3,58
2	0,65	1	0,33	1	0,33	3	0,98	31	10,1
4	1,3	1	0,33	5	1,63	2	0,65	22	7,17
2	0,65	12	3,91	35	11,4	2	0,65	31	10,1

Quadro 49: Agência 06 – Performance utilizando média móvel – experimento 12.

Agência 06									
MM 05		MM 10		MM 15		MM 20		Sem MM	
Acertos	Perc. %	Acertos	Perc. %	Acertos	Perc. %	Acertos	Perc. %	Acertos	Perc. %
52	16,94	38	12,38	45	14,66	37	12,05	23	7,49
41	13,36	45	14,66	43	14,01	31	10,1	20	6,51
53	17,26	47	15,31	44	14,33	30	9,77	25	8,14
54	17,59	46	14,98	36	11,73	35	11,4	31	10,1
54	17,59	42	13,68	37	12,05	34	11,07	22	7,17
50	16,29	37	12,05	36	11,73	43	14,01	25	8,14
47	15,31	55	17,92	45	14,66	36	11,73	31	10,1
55	17,92	52	16,94	30	9,77	32	10,42	18	5,86
55	17,92	43	14,01	43	14,01	30	9,77	21	6,84
49	15,96	39	12,7	38	12,38	36	11,73	21	6,84

Quadro 50: Agência 13 – Performance utilizando média móvel – experimento 12.

Agência 13									
MM 05		MM 10		MM 15		MM 20		Sem MM	
Acertos	Perc. %	Acertos	Perc. %	Acertos	Perc. %	Acertos	Perc. %	Acertos	Perc. %
19	6,19	13	4,23	16	5,21	15	4,89	18	5,86
18	5,86	20	6,51	21	6,84	19	6,19	23	7,49
20	6,51	16	5,21	19	6,19	14	4,56	22	7,17
19	6,19	19	6,19	15	4,89	19	6,19	19	6,19
15	4,89	21	6,84	12	3,91	20	6,51	14	4,56
15	4,89	25	8,14	14	4,56	18	5,86	21	6,84
16	5,21	17	5,54	18	5,86	17	5,54	18	5,86
15	4,89	20	6,51	17	5,54	20	6,51	17	5,54
19	6,19	14	4,56	16	5,21	16	5,21	21	6,84
18	5,86	20	6,51	15	4,89	15	4,89	25	8,14

15. MODELO NEBULOSO

Quadro 51: Ag. 02 – Resultados da aplicação dos modelos fuzzy

Tipo <i>Fuzzificação</i>	Atraso	Teste	
		Acertos (309 reg)	%
1	1	20	6,49
1	2	25	8,14
1	3	25	8,17
1	4	23	7,54
1	5	14	4,61
1	6	9	2,97
1	7	8	2,65
1	8	5	1,66
1	9	6	2,00
1	10	6	2,01
2	1	27	8,77
2	2	35	11,40
2	3	21	6,86
2	4	10	3,28
2	5	8	2,63
2	6	17	5,61
2	7	22	7,28
2	8	20	6,64
2	9	21	7,00
2	10	18	6,02
3	1	6	1,95
3	2	9	2,93
3	3	14	4,58
3	4	11	3,61
3	5	16	5,26
3	6	16	5,28
3	7	13	4,30
3	8	11	3,65
3	9	13	4,33
3	10	12	4,01
4	1	22	7,14
4	2	20	6,51
4	3	12	3,92
4	4	12	3,93
4	5	10	3,29
4	6	10	3,30
4	7	7	2,32
4	8	4	1,33
4	9	6	2,00
4	10	7	2,34

Quadro 52: Ag. 06 – Resultados da aplicação dos modelos fuzzy

Tipo Fuzzificação	Atraso	Teste	
		Acertos (309 reg)	%
1	1	20	6,49
1	2	18	5,86
1	3	19	6,21
1	4	11	3,61
1	5	16	5,26
1	6	5	1,65
1	7	5	1,66
1	8	14	4,65
1	9	19	6,33
1	10	20	6,69
2	1	22	7,14
2	2	23	7,49
2	3	9	2,94
2	4	9	2,95
2	5	12	3,95
2	6	22	7,26
2	7	18	5,96
2	8	16	5,32
2	9	23	7,67
2	10	24	8,03
3	1	16	5,19
3	2	19	6,19
3	3	20	6,54
3	4	13	4,26
3	5	20	6,58
3	6	27	8,91
3	7	20	6,62
3	8	25	8,31
3	9	19	6,33
3	10	17	5,69
4	1	19	6,17
4	2	18	5,86
4	3	19	6,21
4	4	15	4,92
4	5	14	4,61
4	6	8	2,64
4	7	16	5,30
4	8	17	5,65
4	9	16	5,33
4	10	16	5,35

Quadro 53: Ag. 13 – Resultados da aplicação dos modelos fuzzy

Tipo Fuzzificação	Atraso	Teste	
		Acertos (309 reg)	%
1	1	34	11,04
1	2	31	10,10
1	3	21	6,86
1	4	20	6,56
1	5	12	3,95
1	6	8	2,64
1	7	3	0,99
1	8	7	2,33
1	9	8	2,67
1	10	8	2,68
2	1	23	7,47
2	2	13	4,23
2	3	8	2,61
2	4	7	2,30
2	5	16	5,26
2	6	18	5,94
2	7	19	6,29
2	8	20	6,64
2	9	23	7,67
2	10	13	4,35
3	1	11	3,57
3	2	7	2,28
3	3	11	3,59
3	4	14	4,59
3	5	8	2,63
3	6	9	2,97
3	7	11	3,64
3	8	15	4,98
3	9	16	5,33
3	10	13	4,35
4	1	14	4,55
4	2	12	3,91
4	3	8	2,61
4	4	9	2,95
4	5	4	1,32
4	6	4	1,32
4	7	5	1,66
4	8	7	2,33
4	9	5	1,67
4	10	7	2,34

1. REFERÊNCIAS

- ABELÉM, Antonio J. G. **Redes Neurais Artificiais na Previsão de Séries Temporais**. 1994. 108 p. Dissertação (Mestrado em Engenharia Elétrica) – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.
- AFIF, A. A.; AZEN, J. P. **Statistical Analysis – A Computer Oriented Approach**. 2a. ed. New York: Academic Press, Inc. 1979. 442 p.
- BACHA, Hamid; MEYER, Walter. **A Neural Network Architecture for Load Forecasting**. In: IEEE International Joint Conference on Neural Networks, 2, p. 442-447.
- BISHOP, Christopher M. **Neural Networks for Pattern Recognition**. Oxford: Oxford University Press Inc., 1996. 482 p.
- BOWERMAN, Bruce L.; O'CONNELL, Ricchard T. **Forecasting & Time Series**. Belmont: Duxbury Press., 1979. 481 p.
- BOX, George E. P.; JENKINS, Gwilym M.; REINSEL, Gregory C. **Time Analysis Forecasting and Control**. New Jersey: Prentice Hall, 1994. 598 p.
- BRAGA, Antônio P.; CARVALHO, André P.L.; LUDEMIR, Teresa B. **Fundamentos de Redes Neurais Artificiais – 11^a. Escola de Computação**. Rio de Janeiro: Imprinta Gráfica e Editora Ltda., 1998. 246 p.
- BRASIL, Banco Central do. **Meio Circulante - Dinheiro em Circulação**. Disponível em <http://www.bcb.gov.br/mPag.asp?perfil=1&cod=164&codP=163&idioma=P>. Acessado em 27/01/2003.
- BUSSAB, Wilton; MORETTIN, Pedro A. **Métodos Quantitativos – Estatística Básica**. 4. ed. São Paulo: Saraiva S/A Livreiros Editores, 2001. 321 p.
- CAMPOS, Geraldo M. **Estatística Prática para Docentes e Pós-Graduandos**. Programa de Incentivo à Produção de Material Didático do SIAE, Pró-Reitorias de Graduação e Pós-Graduação da USP. Disponível em

http://www.forp.usp.br/restauradora/gmc/gmc_livro/gmc_livro.html. Atualizado em: 09/2000 à 01/2001. Acesso em: 15/06/2002.

CASTELLANO, Giovanna; FANELLI, Anna M. **Variable selection using neural-network models**. In: Neurocomputing 31 (2000) 1-13. Elsevier Science B.V. 2000, 13 p.

CICHOCKI, Andrej; UNBERHAUEN, Rolf. **Neural Networks for Optimization and Signal Processing**. New York: John Wiley & Sons, 1995. 526 p.

EBECKEN, Nelson F. F. **Resumo de Probabilidade e Estatística**. Apostila do curso “Reconhecimento de Padrões”. Rio de Janeiro: COPPE/UFRJ, 2001. 60 p.

EYDEN, Robert J. V. **The Application of Neural Networks in the Forecasting of Share Prices**. New York: Finance & Technology Publishing, 1996. 286 p.

FAHLMAN, Scott E. **An Empirical Study of Learning Speed in Back-Propagation Networks**, Relatório Técnico, CMU-CS-88-162. da Carnegie Mellon University, 1988. Disponível em <http://www.is.cs.cmu.edu/ISL.html>, acesso em 06/2000. 16 p.

FERNANDES, Luiz G. L. **Utilização de Redes Neurais na Análise e Previsão de Séries Temporais**. 1995. 76 p. Dissertação (Mestrado em Ciência da Computação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre.

FERNANDES, Luiz G. L.; NAVAUX, Philippe O.; PORTUGAL, Marcelo S., **Previsão de Séries de Tempo: Redes Neurais Artificiais e Modelos Estruturais**. In: XVII Encontro Brasileiro de Econometria, Salvador, 1996, v.26, n.2, p. 253-276.

HAIR JR., Joseph F.; ANDERSON, Rolph E. et al. **Multivariate Data Analysis** 5. ed. New Jersey: Prentice Hall, 1998. 730 p.

HAN, Jiawei; KAMBER, Micheline **Data Mining: Concepts and Techniques**. San Francisco: Morgan Kaufmann Publishers, 2001. 550 p.

HAYKIN, Simon **Neural Networks - A Comprehensive Foundation** 2. ed. New Jersey: Prentice Hall, 1999. 842 p.

_____. **Redes Neurais – Princípios e Prática**. Porto Alegre: Bookman Companhia Editora, 2001. 900 p.

IATROS. **Estatística e Pesquisa Científica para Profissionais de Saúde**. Estatística Inferencial. Coeficiente de correlação. Disponível em <http://www.vademecum/com.br/iatros/correlação.htm>. Acesso em: 04 abr. 2002.

JAIN, Anil; ZONGKER, Douglas. **Feature Selection: Evaluation, Application, and Small Sample Performance**. Disponível em <http://www.doc.ic.ac.uk/~xh1/Referece/feature-selection/feature-selection>. Acesso em: 16 mar. 2004.

KAPADIA, Apu; LU, Charng-da. **Random Number Generators**. Disponível em <http://devius.cs.uiuc.edu/~akapadia/project2/>. Acesso em: 23/05/2003.

KOHONEN, Teuvo. **State of the Art in Neural Computing**. In: First IEEE International Conference on Neural Networks I-79 I-90. San Diego, 1987.

KOVÁCS, Zsolt L. **Redes Neurais Artificiais: Fundamentos e Aplicações** 2. ed. São Paulo: Edição Acadêmica, 1996. 174 p.

LAWRENCE, Steve; GILES, C. Lee; TSOI, Ah C. **What Size Neural Network Gives Optimal Generalization? Convergence Proprieties of Backpropagation**. Relatório Técnico CS-TR-3617 e UMIACS-TR-96-22, University of Maryland, 1996. 37 p.

LEERINK, Laurens R.; GILES, C. Lee; HORNE, Bill G. et al. **Product Unit Learning**. Relatório Técnico CS-TR-3503 e UMIACS-TR-95-80, University of Maryland, 1995. 17 p.

LEVINE, David M.; BERESON, Mark L.; STEPHAN, David. **Estatística: Teoria e Aplicações Usando o Microsoft Excel em Português**. Rio de Janeiro: LTC - Livros Técnicos e Científicos Editora S.A., 2000. 812 p.

LIMA, Edison C. P. de; SAGRILLO, Luís Volnei Sudati. **Confiabilidade Estrutural**. Rio de Janeiro: Curso MSO – Programa de Engenharia Civil – UFRJ/COPPE. Disponível em <http://www.oceanica.ufrj.br/MSO/downloads/ceoff/>. Acesso em 23/05/2003.

LITTLE, Roderick J. A.; RUBIN, Donald B. **Statistical Analysis with Missing Data**. New York: John Wiley & Sons, 1987. 278 p.

MAREN, Alianna, J.; HARSTON, Craig; PAP, Robert; et al. **Handbook of Neural Computing Applications**. San Diego: Academic Pres, Inc, 1990. 448 p.

MONTGOMERY, Douglas C.; JOHNSON, Lynwood A.; GARDINER, John S. **Forecasting & Time Series Analysis** 2. ed. New York: McGraw-Hill Inc., 1990. 381 p.

MORCILLO, André M. **Teste do qui-quadrado (X^2)**. Disponível em http://www.fcm.unicamp.br/centros/ciped/mp639/teste%20Qui_quadrado.pdf. Acessado em 02/04/2004.

_____. **Teste de Kolmogorov-Smirnov**. Disponível em <http://www.fcm.unicamp.br/centros/ciped/mp639/Teste%20de%20Kolmogorov.pdf>. Acessado em 13/04/2004.

MORETTIN, Pedro A.; TOLOI, Clélia M. de C. **Previsão de Séries Temporais**. 2. ed. São Paulo: Atual Editora Ltda, 1987. 439 p.

NASCIMENTO JR., Cairo L.; YONEYAMA, Takashi. **Inteligência Artificial em Controle e Automação**. 1^a. reimpressão. São Paulo: Editora Edgard Blücher Ltda., 2002. 218 p.

NETER, John; WASSERMAN, William; WHITMORE, G. A.. **Applied Statistics**. 4. ed. Boston: Simon & Schuster, Inc., 1992. 992 p.

NEVES, Chaves das. **The Web Page Chemometrics Text Book**. Universidade Nova Lisboa. Monte da Caparica: Faculdade de Ciências e Tecnologia. Disponível em <http://www.dq.fct.unl.pt/qof/chemidx1.html#index>. Acesso em 23/05/2003.

OLIVA, Gláucia M.; CYBIS, Helena B. B.; PRETTO, Carlos Oliva. **Metodologia de Redes Neurais Aplicada à Previsão de Produção de Viagens**. XV Congresso de Pesquisa e Ensino em Transportes, Campinas. 2001. p. 283-290.

PAIVA, Wagner P. **A Teoria do Caos e as Organizações**. In: Caderno de Pesquisas em Administração vol. 08, no. 2, abril/junho 2001. São Paulo, 2001. p. 85-96.

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. **Learning Representations by Back-Propagating Errors**. In: Nature vol. 323, 1986. p. 533-536.

RUMELHART, David E.; WIDROW, Bernard; LEHR, Michael. **The Basic Ideas in Neural Networks**. In: Communications of the ACM vol. 37, no. 3, 1994. p. 87-92.

SÄRNDAL, Carl-Erik; SWENSSON, Bengt; WRETMAN, Jan. **Model Assisted Survey Sampling – Spring Series in Statistics**. New York: Springer-Verlag Inc. 1992. 694 p.

SILVA, Marco A. V. R. **A Hipótese da Eficiência do Mercado Acionário e Algumas Anomalias Encontradas**. Disponível em <http://www.unitau.br/nupes/hipotese.htm>. Acessado em 19/03/2004.

SINCICH, Terry. **Business Statistics by Example**. 5. ed. New Jersey: Prentice Hall, 1996. 1258 p.

SOARES FILHO, William. **Classificação do Ruído Irrradiado por Navios usando Redes Neurais**. Tese (Doutorado em Engenharia Elétrica) – Coordenação dos Programas de Pós-Graduação em Engenharia – COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro. 2001. 194p.

SPIEGEL, Murray R. **Estatística**. Rio de Janeiro: SEDEGRA – Sociedade Editora e Gráfica Ltda, 1961. 580 p.

STEVENSON, William J. **Estatística Aplicada à Administração**. São Paulo: Harbra Ltda., 1986. 495 p.

STORNETA, W. S.; HUBERMAN, B. A. **An Improved Three-Layer Back-Propagation Algorithm**. In: First IEEE Conference on Neural Networks, II-637 II-644. San Diego. 1990.

TAFNER, Malcon A.; XEREZ, Marcos de; RODRIGUES FILHO, Ilson W. **Redes Neurais Artificiais – Introdução e Princípios de Neurocomputação**. Blumenau: Editora Eko: Editora da FURB, 1995. 199 p.

TAKEMURA, Roberto Yukio. **Controle Inteligente – Tutorial**. Universidade Estadual de Maringá. Disponível em http://www.din.uem.br/ia/control/fuz_prin.htm. Acessado em 15/12/2003.

TENÓRIO, Manoel F. e LEE, Wei-Tsih, **Self Organizing Neural Network for Optimun Supervised Learning**, TR-EE 89, Purdue University, 1989.

THOMÉ, Antonio C. G. “**Notas de aulas**”, Curso de pós-graduação do IM\DCC\NCE – UFRJ (RJ), 2000, 23 p.

_____. **Roteiro para Aplicação do CPD e Relevância das Entradas**, Notas de aulas da pós-graduação do IM\DCC\NCE – UFRJ (RJ), 2001, 2 p.

_____. **Aula 11 – MLP (MultiLayer Perceptron)**, Notas de aulas da pós-graduação do IM\DCC\NCE – UFRJ (RJ), 2003, 37 p.

VIANA, Gizelle K. **Sistemas Híbridos Inteligentes para Reconhecimento de Padrões**. Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro – UFRJ, Rio de Janeiro, 2000, 156 p.

ZADEH, L. A. **Fuzzy Sets**. Information and Control, n.8, 1965, p.338-353.

ZHANG, Guoqiang; PATUWO, B. Eddy; HU, Michael Y. **Forecasting with artificial neural networks: The state of the art**. International Journal of Forecasting. 14 (1998) 35-62.

ZURADA, Jacek M. **Introduction to Artificial Neural Systems**. St. Paul: West Publishing Company, 1992. 683 p.