

INVESTIGAÇÃO DE TÉCNICAS DE SIMULAÇÃO NA MODELAGEM DE REDE  
NEURAL ARTIFICIAL APLICADA A TRANSPLANTE RENAL

João Carlos da Gama Dias Costa

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS  
PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE  
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS  
PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA  
BIOMÉDICA.

Aprovada por:

---

Prof. Antonio Fernando Catelli Infantsi, Ph.D.

---

Prof. Renan Moritz Varnier Rodrigues de Almeida, Ph.D.

---

Prof. Jurandir Nadal, D.Sc.

---

Prof. José Hermógenes Rocco Suassuna, D.Sc.

---

Prof. Marcelo Cunha Medeiros, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

ABRIL DE 2005

COSTA, JOÃO CARLOS DA GAMA DIAS

Investigação de Técnicas de Simulação na  
Modelagem de Rede Neural Artificial Aplicada a  
Transplante Renal [Rio de Janeiro]

2005

XI, 115 p. 29,7 cm (COPPE/UFRJ, M.Sc.,  
Engenharia Biomédica, 2005)

Tese - Universidade Federal do Rio de  
Janeiro, COPPE.

1. Redes Neurais Artificiais
2. *Bootstrap*
3. Transplante Renal
4. Modelagem Matemática

I. COPPE/UFRJ II. Título ( série )

"O valor das coisas não está no tempo em que elas duram, mas na intensidade com que acontecem. Por isso, existem momentos inesquecíveis, coisas inexplicáveis e pessoas incomparáveis".

(Fernando Pessoa)

“Saúdo-te, Esperança, tu que vens de longe,  
inundas com teu canto os tristes corações,  
tu que dá novas asas aos sonhos mais antigos,  
tu que nos enche a alma de brancas ilusões.

Saúdo-te, Esperança. Tu forjarás os sonhos  
naquelas solitárias desenganadas vidas,  
carentes do possível de um futuro risonho,  
naquelas que inda sangram as recentes feridas.

Ao teu sopro divino fugirão as dores  
como tímido bando de ninho despojado,  
e uma aurora radiante, com suas belas cores,  
anunciará às almas que o amor é chegado.”

(Pablo Neruda)

## DEDICATÓRIA

A meus pais e irmãos, família,  
porto seguro de todas as horas.

A você, Vívian, amor maior, que  
desde seu nascimento naquele  
lindo anoitecer em Belo Horizonte,  
iluminou definitivamente a minha  
vida.

## AGRADECIMENTOS

Aos professores Fernando Infantsi e Renan Moritz por terem me ajudado e orientado por todo o tempo do mestrado e, principalmente, de confecção da dissertação. Suas contribuições ao tema foram enormes e, claro, os erros contidos no trabalho são de minha exclusiva responsabilidade. Também aos professores Giannella Neto e Frederico Jandre, por terem compreendido que uma tese deve ser feita com paixão e este tema me era tão caro. Ao professor Márcio Souza, que me aceitou junto à sua turma no Doutorado, de Identificação de Sistemas, e contribuiu de forma decisiva para que me interessasse pelo assunto “Redes Neurais”. Aos demais professores do PEB, pelo convívio enriquecedor, ensinamentos e amizade. Aos colegas Amauri, Diniz, Luciano e Luis, pelas horas de conversa descontraída e ajuda nas suas diversas especialidades. E, claro, às meninas Anacleide e Daisy, pela ajuda nos assuntos administrativos e pelas dicas sempre valiosas na hora das matrículas, registro etc..

Os colegas que encontrei no PEB são um caso à parte, pela juventude, vontade de crescer e camaradagem própria dos ambientes acadêmicos sadios. Fizeram-me lembrar dos tempos de faculdade, já um pouco mais longe. Sem dúvida, lhes sou devedor pela experiência maravilhosa deste convívio.

Aos amigos da CEMIG, não vou citar alguém, são tantos, mas espero que saibam o quanto apreciei o tempo em que passei junto a eles, nas Minas Gerais.

Deus colocou anjos no meu caminho, que se tornaram amigos de jornada. Dentre tantos, e para não cometer a injustiça do esquecimento momentâneo, cito apenas aqueles que foram, estão, mais presentes: Dra Deise, Dr. Salomão e Dr. Cyrillo. Que tenham bastante saúde e paz para cuidar de seus pupilos.

A meu pai, que me deu a vida duas vezes. Procuro sempre não desperdiçar a oportunidade. Minha mãe e irmãos pela dedicação, incentivo e criação de condições para que todos atingíssemos nossos objetivos.

A você, Cristina, onde amor, companheirismo, amizade, são algumas das qualidades únicas que nos une. Foram horas, dias, madrugadas roubadas do seu convívio, mas, como você sabe, no fim dá tudo certo.

Acho que é só. Quem não foi citado, por puro esquecimento neste momento, sinta-se, entretanto, lembrado com carinho.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## INVESTIGAÇÃO DE TÉCNICAS DE SIMULAÇÃO NA MODELAGEM DE REDE NEURAL ARTIFICIAL APLICADA A TRANSPLANTE RENAL

João Carlos da Gama Dias Costa

Abril/2005

Orientadores: Antonio Fernando Catelli Infantsi  
Renan Moritz Varnier Rodrigues de Almeida

Programa: Engenharia Biomédica

O transplante renal é o melhor tratamento para a insuficiência renal crônica terminal, porém a sobrevida do enxerto a longo prazo é ainda um problema não resolvido. Neste contexto, a modelagem por Regressão Linear Múltipla (RLM) e um modelo não-linear, baseado em Redes Neurais Artificiais, foram investigados com vistas à predição de sobrevida do enxerto em escala contínua. Este estudo, devido a incompletude das variáveis, baseou-se em dados de somente 109 transplantados com doadores vivos, selecionados retrospectivamente após a introdução da Ciclosporina A como medicamento imunossupressor. Com base na “mineração” dos dados, foram selecionadas 11 variáveis preditoras, tais como Compatibilidade HLA, Tempo de Diálise e Idade do Receptor. O modelo de RLM resultou em percentual de acerto de 6,2% e Coeficiente de Determinação de 9%. Por outro lado, a RNA, treinada com algoritmo de retropropagação e neurônios sigmoidais na predição de sobrevida, apresentou desempenho superior, com percentual de acertos de 34,4% e Coeficiente de Determinação de 47%. Devido ao número limitado de pacientes disponíveis para o estudo, foi desenvolvida uma metodologia baseada em *Bootstrap* e Validação Cruzada e testes estatísticos não-paramétricos, para a seleção e validação da topologia da RNA, resultando em uma estrutura de sete neurônios na camada oculta. Os resultados alcançados e a metodologia proposta indicam a potencialidade da modelagem por RNA na predição da sobrevida do enxerto renal em escala contínua e, portanto, se constituem em contribuição ao desenvolvimento de metodologias para o prognóstico destes pacientes.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

INVESTIGATION OF SIMULATION TECHNIQUES FOR ARTIFICIAL NEURAL  
NETWORK MODELLING APPLIED TO RENAL TRANSPLANTATION

João Carlos da Gama Dias Costa

April/2005

Advisors: Antonio Fernando Catelli Infantosi  
Renan Moritz Varnier Rodrigues de Almeida

Department: Biomedical Engineering Program

Renal transplantation is the best treatment for end-stage renal disease but long-term graft survival is still a problem. In this context, a Multiple Linear Regression model and a non-linear modelling using an Artificial Neural Networks were carried out to predict kidney allograft outcome in a continuous scale. Due to missing data, this work was based on 109 living related renal transplant recipients, retrospectively selected after Cyclosporine A introduction for immunosuppression therapy. Using a data mining approach, 11 predictors, such as HLA Mismatch, Recipient Age and Duration of Dialysis, were determined. The Multiple Linear Regression model resulted in an accuracy of 6.2% and a determination coefficient of 9%. On the other hand, the Neural Network, trained with the backpropagation algorithm and sigmoid neurons to allograft outcome prediction, achieved a better performance, that is, 34.4% of accuracy and determination coefficient of 47%. Due to the small data set available, a methodology using Bootstrap, Cross Validation and nonparametric statistical tests was developed for determining the RNA topology with seven hidden nodes. Both this methodology and the findings achieved by using the RNA indicate the potentiality of modeling the allograft outcome prediction on continuous scale from small data sets. Although further work should be carried out, the RNA seems to constitute an important contribution to prognosis of these patients.

**INVESTIGAÇÃO DE TÉCNICAS DE SIMULAÇÃO NA MODELAGEM DE REDE  
NEURAL ARTIFICIAL APLICADA A TRANSPLANTE RENAL**

**Índice**

<b>CAPÍTULO I – Introdução.....</b>	<b>1</b>
I.1 – Introdução.....	1
I.2 – Objetivo.....	3
I.3 – Estrutura do Trabalho.....	3
<b>CAPÍTULO II - Redes Neurais Artificiais Aplicadas a Transplantes Renal.....</b>	<b>5</b>
II.1 – Introdução.....	5
II.2 – Tipos de Redes Neurais Artificiais (RNA).....	6
II.2.1 - <i>Perceptrons</i> Unicamadas.....	6
II.2.2 - ADALINE e MADALINE.....	7
II.2.3 - <i>Perceptrons</i> Multi-Camadas.....	9
II.3 – Técnicas de Reamostragem.....	11
II.4 – Revisão Bibliográfica sobre Utilização de RNA em Transplantes.....	12
<b>CAPÍTULO III – Análise Exploratória de Dados de Transplantados Renais.....</b>	<b>25</b>
III.1 – Introdução.....	25
III.2 – Definição das Variáveis Dependente e Predictoras.....	25
III.2.1 – Variáveis Contínuas.....	27
III.2.2 – Variáveis Dicotômicas.....	31
III.2.3 – Variáveis Categóricas.....	31
III.3 – Padronização da Variáveis.....	32
III.4 – Conjuntos de Dados.....	32
III.5 – Discussão.....	35
<b>CAPÍTULO IV - Modelagem Linear Multivariável Aplicada a Transplante Renal.....</b>	<b>37</b>
<b>IV.1 – Introdução.....</b>	<b>37</b>
IV.2 – Regressão Linear Múltipla (RLM) pelo Método dos Mínimos Quadrados.....	37
IV.3 – Índices de Desempenho.....	39
IV.4 – Aplicação da RLM em Transplante Renal.....	40
IV.5 – Considerações Sobre a Utilização de RLM em Transplantes.....	43
<b>CAPÍTULO V - Modelagem por RNA em Transplante Renal.....</b>	<b>46</b>
V.1 – Introdução.....	46
V.2 – Definição do número de Replicações <i>Bootstrap</i> (BST).....	46
V.3 – Aplicação de RNA em Transplante Renal.....	49

V.3.1 – Inicialização de Pesos.....	49
V.3.2 – Seleção da Arquitetura da RNA.....	49
V.3.2.1 - Determinação do Número de Neurônios Ocultos.....	50
V.3.2.2 - Eliminação das Arquiteturas com Menor Proporção de Acertos para o Conjunto de Estimação .....	52
V.3.2.3 - Verificação da Proporção de Acertos com e sem BST.....	51
V.3.3 – Verificação da Diferença Entre Arquiteturas de Menor Quantidade de Parâmetros Com e Sem BST.....	55
V.4 – Verificação da generalização.....	57
V.5 – Desempenho da Rede Neural Artificial Seleccionada.....	58
V.6 – Discussão.....	60
<b>CAPÍTULO VI – Conclusão.....</b>	<b>65</b>
<b>CAPÍTULO VII – Referências Bibliográficas.....</b>	<b>68</b>
<b>APÊNDICE A – Tópicos em Redes Neurais Artificiais.....</b>	<b>78</b>
A.1 – Introdução.....	78
A.2 – Regras de Aprendizagem por Correção do Erro.....	81
A.3 – Técnicas de Otimização – Gradiente Descendente.....	82
A.4 – Algoritmo de Retropropagação do Erro (BP).....	85
<b>APÊNDICE B – Características das Variáveis de Transplante Renal .....</b>	<b>91</b>
B.1 – Relação dos Pacientes (Receptores).....	91
B.2 – Variáveis Dependente e Preditoras.....	92
B.3 – Variáveis Padronizadas.....	97
B.4 – Conjuntos de Estimação, Validação e Teste.....	104
<b>APÊNDICE C – <i>Bootstrap</i> (BST).....</b>	<b>108</b>
C.1– Introdução.....	108
C.2 – Validação do LabVIEW®.....	108
C.3 – Algumas Aplicações do BST.....	112

## LISTA DE ABREVIATURAS

**ACE** – Proporção de acertos no Conjunto de Estimação;  
**ACT** – Proporção de acertos no Conjunto de Teste;  
**ACV** – Proporção de acertos no Conjunto de Validação;  
**BP** – Algoritmo de Retropropagação do Erro;  
**BST** – *Bootstrap*;  
**BSTbal** – *Bootstrap* balanceado;  
**CAPD** – *Continuous Ambulatory Peritoneal Dialysis* (Diálise Peritonial);  
**CMV** – Citomegalovírus;  
**CsA** – Ciclosporina A;  
**EI** – Estatística de Interesse;  
**EQM** – Erro Quadrático Médio;  
**FA** – Função de Ativação do Neurônio Artificial;  
**FO** – Função Objetivo;  
**GD** – Gradiente Descendente;  
**GDC** – Gradiente Descendente Conjugado;  
**H<sub>0</sub>** – Hipótese Nula  
**HCV** – Vírus da Hepatite C;  
**HD** – Hemodiálise;  
**HLA** – *Human Leukocyte Antigen*;  
**MAR** – *Missing At Random*;  
**MCAR** – *Missing Completely At Random*;  
**MDM** – *Missing Data Mechanism*;  
**MMQ** – Método dos Mínimos Quadrados;  
**MNAR** – *Missing Not At Random*;  
**MSE** – *Mean Square Error*;  
**PMC** – *Perceptron* Multi-Camada;  
**PRA** – Painel de Reatividade;  
**PUC** – *Perceptron* Unicamada;  
**RL** – Regressão Logística;  
**RLM** – Regressão Linear Múltipla;  
**RMSE** – *Root Mean Square Error*;  
**RNA** – Rede Neural Artificial;

**ROC** – *Receiver Operating Characteristic*;

**SE** – Sobrevida do Enxerto;

**TX** – Transplante;

**TXH** – Transplante Hepático;

**TXP** – Transplante Pancreático;

**TXR** – Transplante Renal;

# CAPÍTULO I

## INTRODUÇÃO

### I.1- Introdução

A prevalência da insuficiência renal crônica terminal no Brasil é de aproximadamente 300 pacientes por milhão de população por ano, com 17.000 pacientes começando tratamento dialítico a cada ano (GARCIA *et al.*, 2003). Mesmo com a evolução das terapias renais substitutivas, como hemodiálise, diálise peritoneal e transplante, a expectativa de vida nestes pacientes ainda é menor do que em outros grupos populacionais de mesma faixa etária, sendo particularmente afetada em pacientes diabéticos, de raça branca e em idosos (LOPES *et al.*, 2003). Problemas cardiovasculares, anormalidades no sistema de coagulação e distúrbios nutricionais não são incomuns em pacientes urêmicos e em transplantados renais (SALOMÃO-FILHO *et al.*, 1996; VERPOOTEN *et al.*, 1996; SARTORI *et al.*, 2000; GUARNIERI *et al.*, 2001; RIELA e MARTINS, 2001; PERCOITS-FILHO *et al.*, 2002; PERKOWSKA *et al.*, 2002; MANRIQUE *et al.*, 2003; RADERMACHER *et al.*, 2003; ROBINSON, 2004), o que contribui para a diminuição da sobrevida dos pacientes. O Brasil é o 4º país em número de transplantes renais, após somente dos EUA, China e Índia. O custo com tratamento dialítico coberto pelo SUS é de cerca de 300 milhões de dólares por ano, enquanto que para transplante renal é de aproximadamente 100 milhões de dólares por ano (GARCIA *et al.*, 2003).

O transplante renal (TXR) bem sucedido é o melhor tratamento disponível para a insuficiência renal crônica terminal (SALOMÃO-FILHO *et al.*, 1996; Ávila *et al.*, 2004), embora tenha havido importantes avanços no controle da perda do enxerto no primeiro ano pós-transplante, principalmente devido às novas drogas imunossupressoras, o desafio da sobrevida do enxerto a longo prazo ainda persiste (MATSUI *et al.*, 2001; MARSDEN, 2003). A disfunção crônica do enxerto (ou rejeição crônica) é atualmente a principal causa de perda tardia da função renal em transplantados renais (SIMIC-OGRIZOVIC *et al.*, 1999; MATSUI *et al.*, 2001).

Mais recentemente, as Redes Neurais Artificiais (RNA) têm sido empregadas no auxílio ao diagnóstico de complicações e prognóstico da sobrevida em transplante (renal, pancreático e hepático). DVORCHIK *et al.* (1996) estudaram a aplicabilidade do emprego de Redes Neurais Artificiais na predição da sobrevida de enxerto hepático,

enquanto HSIANG-YIN *et al.* (1999), investigaram a predição de níveis séricos de imunossupressor em transplantados hepáticos. MELVIN *et al.* (2000) e HUGHES *et al.* (2001) estudaram a aplicação de RNA no diagnóstico de rejeição em transplante hepático, enquanto DORSEY *et al.* (1997) aplicaram na predição da sobrevida em transplante pancreático.

As Redes Neurais Artificiais também foram aplicadas à predição do funcionamento do enxerto renal (SHOSKES *et al.*, 1998); ao diagnóstico de rejeição em transplante renal (FURNESS *et al.*, 1999), na seleção de variáveis mais significativas na predição de rejeição crônica de transplante renal (SIMIC-OGRIZOVIC *et al.*, 1999), na determinação de citomegalovirose em receptores de TXR (SHEPPARD *et al.*, 1999) e na dosagem de imunossupressor em transplante renal (CAMPS-VALLS *et al.*, 2003). Recentemente, CROSS *et al.* (2003) enfatizaram o uso da RNA na clínica médica, reforçando que a aceitabilidade de sistemas computacionais de suporte à decisão, integrados à estrutura existente, deverá aumentar. Estes trabalhos mostram a viabilidade do uso de Redes Neurais Artificiais, em comparação com os métodos de modelagem estatística tradicionais.

Uma das limitações da modelagem estatística em saúde deve-se ao número pequeno, em geral insuficiente, de casos disponíveis para análise. Esta limitação dificulta sobremaneira o estabelecimento de pressupostos (hipóteses) necessários ao desenvolvimento de um modelo linear, por exemplo, Gaussianidade e homocedasticidade, sugerindo o uso de modelos não-lineares, tais como as Redes Neurais Artificiais (SHOSKES *et al.*, 1998; GISMONDI, 1999; ALMEIDA *et al.*, 2001; INFANTOSI *et al.*, 2004). Outra limitação refere-se à existência de dados perdidos em variáveis de interesse ou consideradas mais relevantes, que, muitas vezes, se observa quando se analisam retrospectivamente dados de pacientes (CLARK e ALTMAN, 2003). A estratégia usualmente empregada para lidar com este problema tem sido o descarte de casos cujas variáveis estejam incompletas (DVORCHIK *et al.*, 1996; SHEPPARD *et al.*, 1999; CLARK e ALTMAN, 2003; GOLDFARB-RUMYANTZEV *et al.*, 2003), com implicações na capacidade do modelo preditivo, ou seja, menor acurácia e resultados tendenciosos, em particular no cenário de doenças de baixa prevalência (OOSTENBRINK *et al.*, 2003).

No contexto das RNA, HAYKIN (2002) recomenda que os exemplos de casos apresentados ao treinamento da RNA sejam os mais representativos possíveis do cenário de interesse, e que o aumento do conteúdo de informação seja alcançado pela diversidade de exemplos usados neste treinamento da RNA, de modo a possibilitar

maior generalização dos resultados. Assim, a reamostragem dos dados de treinamento por *Bootstrap* (método computacionalmente intensivo, baseado na amostragem com reposição dos dados originais) poderia ser empregada para o aumento da informação da base de dados original, tornando mais realista a análise dos resultados do modelo (ZHANG, 1999; HAHN *et al.*, 2005), e sendo, portanto, útil em modelagens com pequeno tamanho amostral.

A predição dicotômica de sobrevida em transplante renal tem sido estudada, indicando a viabilidade do uso de Redes Neurais Artificiais nesta área. Entretanto, pelo nosso conhecimento, não há publicação utilizando RNA na sobrevida em escala contínua, pois o problema é complexo (SHEPPARD *et al.*, 1999). A busca por soluções, que auxiliem a decisão em questões clínicas relativas a transplante, poderia ajudar na identificação de fatores que influenciam na sobrevida do enxerto e em ações corretivas (por exemplo, alterando a medicação imunossupressora) (CROSS *et al.*, 2003; GOLDFARB-RUMYANTZEV *et al.*, 2003).

## **I.2- Objetivo**

Determinar metodologias para seleção da topologia e validação de uma Rede Neural Artificial na predição de sobrevida em escala contínua (meses) de enxerto renal em receptores de doadores vivos, usando *Perceptron* Multi-Camada (PMC) com o algoritmo de retropropagação do erro e técnicas de computação intensivas como o *Bootstrap* (BST) e Validação Cruzada.

## **I.3- Estrutura do Trabalho**

O presente trabalho foi desenvolvido em tópicos, com a introdução às Redes Neurais Artificiais no Capítulo II, ressaltando a característica de “Aproximador Universal de Funções” das redes, a metodologia BST utilizada no trabalho e a revisão bibliográfica dos artigos encontrados utilizando RNA em transplantes, não somente renais. A “mineração” das variáveis é mostrada no Capítulo III, enquanto a metodologia da análise multivariável por Regressão Linear Múltipla está desenvolvida no Capítulo IV. No Capítulo V são mostrados os critérios de estabelecimento da topologia da rede, a definição do número de replicações *Bootstrap*, a comparação com o modelo de Regressão Linear Múltipla e a avaliação dos resultados. A conclusão está no Capítulo VI, e as referências bibliográficas utilizadas estão listadas no Capítulo VII.

No Apêndice A é feita uma introdução ao algoritmo de retropropagação de erro, no Apêndice B são descritas as variáveis e feitas observações sobre a base de dados, enquanto no Apêndice C é introduzida a teoria *Bootstrap*.

## CAPÍTULO II

### Redes Neurais Artificiais Aplicadas a Transplante Renal

#### II.1) Introdução

As Redes Neurais Artificiais (RNA) são estruturas de modelagem tipo “Caixa-Preta” (LJUNG, 1999; AGUIRRE, 2000), ou “Aproximadoras de Funções”, utilizadas para classificação, identificação e predição de sistemas lineares e não-lineares. Tentam modelar o comportamento do neurônio humano, ou de seu conjunto, chamado por AZEVEDO *et al.* (2000) de Rede Neural Natural (RNN), utilizando variações nas suas conexões através de parâmetros denominados de pesos sinápticos (doravante intitulados somente pesos). Assim como um agrupamento neuronal, no qual várias sinapses de entrada de um dado neurônio contribuem para ativação (aparecimento do potencial de ação) ou inibição (repouso) de outros neurônios (GUYTON e HALL, 1998), as entradas da Rede Neural Artificial também possuem estas características. Como os dados de entrada já são inerentes ao sistema a ser estudado, a ativação ou inibição de um neurônio é devida, principalmente, ao peso dado a esta conexão. Se positivo, o estímulo é excitatório, e, se negativo, é inibitório (GISMONDI, 1999; de AZEVEDO *et al.*, 2000; HAYKIN, 2002). A “força” da conexão sináptica está associada ao módulo do peso relativo àquela conexão.

A seleção de modelos de Redes Neurais Artificiais (entendendo-se aqui como incluindo arquitetura e pesos) pode demandar grande tempo computacional para o treinamento da rede, sendo que várias técnicas têm sido propostas para diminuir este tempo (HAYKIN, 2002). A abordagem na seleção dos modelos de RNA neste trabalho incluiu técnicas de reamostragem, como o *Bootstrap*. Neste caso, a hipótese assumida ao se reamostrar os dados do conjunto de treinamento é de que haveria a geração de novas amostras com mesmas propriedades estatísticas do conjunto original e, portanto, propiciando a escolha de um modelo com melhor generalização (HAHN *et al.*, 2005).

No contexto da aplicação de RNA como ferramenta de auxílio a decisões médicas em Transplantes (TX), não há muitos artigos publicados, principalmente na predição de sobrevida em TX, apesar do avanço da utilização desta metodologia na área médica (HUGHES *et al.*, 2001). Por esse motivo, será feita aqui uma revisão de artigos associando RNA e TX para problemas similares.

## II.2 – Tipos de RNA

As RNA possuem topologias (arquiteturas), regras de aprendizagem, algoritmos para minimização da função objetivo (ou função custo) de acordo com o ambiente em que estão inseridas, ou seja, de acordo com os dados que fazem parte da matriz de entrada e saída (composta de vetores de entrada e saída dos casos a serem apresentados à RNA). Dentre os vários tipos de Redes Neurais Artificiais, serão introduzidos aqui o *Perceptron* Unicamada, as redes ADALINE e MADALINE e o *Perceptron* Multi-Camada.

### II.2.1 *Perceptrons* Unicamadas (PUC)

O PUC foi desenvolvido por Rosenblat, inicialmente para o reconhecimento de caracteres (GISMONDI, 1998; de AZEVEDO, 2000, HAYKIN, 2002). Este é basicamente uma rede neural classificadora de padrões, para a qual a convergência ocorre desde que o vetor de entrada possa ser enquadrado em classes linearmente separáveis por um hiperplano definido por (HAYKIN, 2002):

$$\mathbf{W}^T \cdot \mathbf{X} + b = 0 \quad (\text{II.1})$$

Um hiperplano é um subconjunto do espaço vetorial  $\mathfrak{R}^m$  e o divide em dois sub-espaços (BOLDRINI *et al.*, 1982; TAJINE e ELIZONDO, 1998). Assim, por exemplo, para o espaço vetorial  $\mathfrak{R}^2$  com vetor de entrada  $\mathbf{X} = (x_1, x_2)$  a equação do hiperplano (neste caso, uma reta) que estabelece uma fronteira de decisão para a RNA pode ser expressa por:

$$w_1 \cdot x_1 + w_2 \cdot x_2 + b = 0 \quad (\text{II.2})$$

onde,

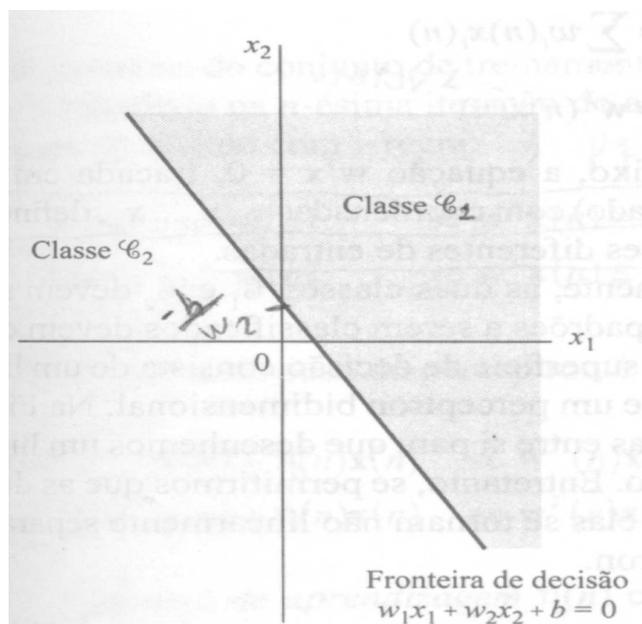
$w_1$  é o coeficiente angular relativo a  $x_1$ ;

$w_2$  é o coeficiente angular relativo a  $x_2$ ;

$b$  é a constante de deslocamento do hiperplano em relação à origem.

No contexto das RNA,  $w_1$  e  $w_2$  são os pesos relativos às entradas  $x_1$  e  $x_2$ , respectivamente,  $b$  é o *bias* e, conforme ilustrado na Figura II.1, o semiplano onde

(II.2) é maior do que zero tem-se a classe  $C_1$ , a qual está associada à resposta +1. Se (II.2) é menor do que zero, a classe é  $C_2$ , e pode ser associada a resposta -1.



**Figura II.1** - Gráfico de um hiperplano dividindo espaço vetorial de 2 dimensões. Os pesos  $w_1$  e  $w_2$  determinam a direção do hiperplano e o *bias*  $b$  o quanto o hiperplano afasta-se da origem. O ponto dado por  $-b / w_2$  é o intercepto da coordenada  $x_2$ . Modificado de HAYKIN, 2002.

### II.2.2 ADALINE e MADALINE

A rede ADALINE (do Acrônimo ADaptive Linear NEuron), introduzida por Widrow, constitui uma variante do PUC. A saída da ADALINE pode assumir valores (+1) ou (-1), dependendo dos valores das entradas. A diferença primordial da ADALINE para o PUC é o tratamento matemático dado ao algoritmo de treinamento, conhecido pelo nome de Regra Delta ou Regra de Aprendizado de Widrow-Hoff, em homenagem aos seus idealizadores, e a utilização da função de ativação linear (de AZEVEDO *et al.*, 2000; HAYKIN, 2002). A Regra Delta utiliza uma Função Objetivo (FO) que é a soma dos erros quadráticos entre a saída desejada do sistema ( $d_i(n)$ ) e a saída da RNA ( $y_i(n)$ ). Assim, o Erro Quadrático (EQ), na camada de saída  $k$ , é dado por:

$$EQ = (d_k(n) - y_k(n))^2 = e_k^2(n) \quad (II.3)$$

onde

$d_k(n)$  é a saída desejada do sistema, na iteração (n);

$y_k(n)$  é a saída da RNA, na iteração (n);

$e_k(n)$  é o erro absoluto entre a saída desejada e a saída da RNA, na iteração (n).

O Erro Quadrático Médio (EQM) ou MSE (*Mean Square Error*) é dado então pela diferença quadrática média entre a saída do sistema e a saída da RNA:

$$EQM(n) = \frac{(d_k(n) - y_k(n))^2}{2} \quad (II.4)$$

A determinação dos pesos é dada pela minimização da Função Objetivo por um método baseado em gradiente, no caso, pelo método do Gradiente Descendente (GD). O GD procura minimizar a superfície do Erro Quadrático Médio no espaço vetorial formado pelos pesos, fazendo com que a parcela de alteração do peso seja proporcional à derivada do EQM em relação a este peso. Então, supondo que a informação propague da camada k para a camada j, a determinação do pesos  $w_{kj}$  é

dada por:

$$w_{kj}(n) = w_{kj}(n-1) + \Delta w_{kj}(n-1) \quad (II.5)$$

onde

$w_{kj}(n)$  é o peso correspondente aos neurônios da camada j vindo da camada k, na iteração (n);

$w_{kj}(n-1)$  é o peso correspondente aos neurônios da camada j vindo da camada k, na iteração (n-1);

$\Delta w_{kj}(n-1)$  é a variação do peso correspondente aos neurônios da camada j vindo da camada k, na iteração (n-1);

e, omitindo-se os índices das iterações:

$$\Delta w_{kj} = -\eta \cdot \frac{\partial EQM}{\partial w_{kj}} \quad (II.6)$$

Mas, pela regra da cadeia do cálculo (LEITHOLD, 1977):

$$\frac{\partial EQM}{\partial w_{kj}} = \frac{\partial EQM}{\partial y_k} \cdot \frac{\partial y_k}{\partial w_{kj}}$$

$$\frac{\partial EQM}{\partial y_k} = -(d_k - y_k) \quad (11.7)$$

$$\frac{\partial y_k}{\partial w_{kj}} = \frac{\partial \sum w_{kj} \cdot u_j + bias_k}{\partial w_{kj}} = u_j$$

Substituindo,

$$\Delta w_{kj} = \eta \cdot (d_k - y_k) \cdot u_j \quad (11.8)$$

ou

$$\Delta w_{kj} = \eta \cdot \delta \cdot u_j \quad (11.9)$$

Logo, a variação no peso pode ser perfeitamente determinada, desde que se conheça a taxa de aprendizagem, o erro entre a saída do sistema e a saída da RNA, e, ainda, a entrada associada àquele peso. A rede MADALINE (*Multiple ADALINE*) é a generalização multidimensional da ADALINE (KOVÁCS, 2002).

### 11.2.3 Perceptrons Multi-Camadas (PMC)

O desenvolvimento do Algoritmo de Retropropagação do Erro (BP - *Back-Propagation*) por Rummelhart, Hilton e Williams em 1986 representou um marco no estudo das redes neurais, pois possibilitou um método de treinamento eficiente do *Perceptron* Multi-Camada (HAYKIN, 2002, KOVÁCS, 2002). O PMC possui as mesmas camadas de entrada e saída que o PUC, porém acrescido de uma ou mais camadas intermediárias (ocultas). A utilização do PMC com o algoritmo BP possibilitou a resolução de problemas não linearmente separáveis (NLS). A apresentação à rede dos dados de entrada e de saída para a mesma iteração é chamada de par entrada-saída (ES) e a apresentação de todos os pares ES de treinamento constitui uma época (HAYKIN, 2002). A maneira como os pares ES são apresentados ao PMC podem ser de dois tipos, o treinamento por lote e o treinamento seqüencial. No treinamento por lote, a atualização das sinapses é feita após todos os pares de treinamento serem apresentados, e no treinamento seqüencial, a atualização é feita após a apresentação de cada par ES de treinamento. Entre a apresentação de cada

par, os resultados são acumulados e é gerada a matriz de pesos e bias da RNA. O Apêndice A faz uma demonstração do algoritmo BP.

Uma questão prática que surge no dimensionamento do PMC é quanto ao número ideal de neurônios na(s) camada(s) oculta(s). De uma maneira geral não há uma regra para o dimensionamento de tais camadas. REED (1993) e HAYKIN (2002) discorrem sobre os métodos de “Poda” (eliminação de conexões / parâmetros até a obtenção de arquitetura com mesmo desempenho) e crescimento da rede; ANDREA e KALAYEH (1991) propuseram uma faixa ótima da razão entradas / neurônios ocultos entre 1,8 e 2,2, para neurônios com função de ativação sigmoideal; DORSEY *et al.* (1997) utilizaram RNA com 20 neurônios na camada oculta com função de ativação Gaussiana e 12 variáveis de entrada; já SHEPPARD *et al.* (1999), com neurônios utilizando função sigmoideal, obtiveram desempenho melhor com uma rede 17x10x2, valores parecidos aos de ANDREA e KALAYEH (1991). Ou seja, similarmente à determinação de qualquer estrutura de um sistema a ser modelado, onde o número de parâmetros depende da complexidade do problema, a determinação do número ótimo de neurônios ocultos da RNA não é simples.

Para teste e validação de uma Rede Neural Artificial, uma vez que se possui um único conjunto de dados, é comum a divisão aleatória do conjunto inicial em dois sub-conjuntos, o Conjunto de Treinamento e o Conjunto de Teste (CT). O Conjunto de Treinamento é adicionalmente dividido nos Conjuntos de Estimção (CE) e de Validação (CV), e CV é usado para selecionar o modelo usando pares ES distintos de CE (HAYKIN, 2002). A validação interna dos dados é medida sobre CT. Este método é considerado o mais simples dos métodos de Validação Cruzada (*Split Sample Cross Validation* ou *Hold-Out Cross Validation*). Alguns autores fazem distinção entre *Split Sample Cross Validation* e *Split Sample Validation* (TERRIN *et al.*, 2003). Neste trabalho foi mantida a nomenclatura de HAYKIN (2002), que o denomina como “Método de Resistência”. Outros métodos conhecidos de Validação Cruzada são a Validação Cruzada Múltipla e o Método Deixe-Um-De-Fora. Estes métodos podem ser utilizados para escolha de modelos, e têm como principal vantagem o fato de que a comparação entre modelos independe de argumentos probabilísticos (LJUNG, 1999; HAYKIN, 2002). A proporção sugerida de pares ES para o Conjunto de Estimção e de Validação é de 80% e 20% do Conjunto de Treinamento, respectivamente (HAYKIN, 2002).

### II.3 - Técnicas de Reamostragem

Em situações nas quais o número de elementos de uma amostra é pequeno e a distribuição estatística é desconhecida, métodos computacionalmente intensivos podem ser utilizados para estimação e inferência sobre a população, permitindo, assim, aumentar o poder do teste estatístico ou a acurácia dos estimadores. Uma destas técnicas é a Permutação (PERM), que resulta do reordenamento aleatório dos elementos desta amostra (podendo seguir algum padrão ou modelo), a partir da qual se determina a Estatística de Interesse (EI) ou se estima a função densidade de probabilidade (fdp). Quando somente a ordem é alterada e simulam-se amostras sem reposição, a escolha da EI ou da Hipótese Nula ( $H_0$ ) torna-se crucial para o problema, pois a média e variância, por exemplo, não se alterarão na simulação. As amostras podem também ser geradas por reamostragem com reposição, podendo-se então fazer inferências sobre quaisquer EI, como no caso do método *Bootstrap* (BST). No contexto das RNA com pequeno número de pares Entrada-Saída, a metodologia BST pode possibilitar um efeito equivalente ao aumento da informação contida no conjunto de treinamento. As técnicas de reamostragem PERM e BST podem ser utilizadas também na análise de séries temporais (Apêndice C).

A metodologia *Bootstrap* consiste em um método de reamostragem com reposição no qual os valores encontrados em uma amostra aleatória de uma população permite estimar a distribuição da própria população (MANLY, 1997).

A partir do trabalho de Efron, professor da Universidade de *Stanford*, Califórnia, em 1979 (MANLY, 1997; DAVISON e HINKLEY, 1997; EFRON, 2000), o *Bootstrap* foi idealizado como uma ferramenta semelhante ao PERM (DAVINSON e HINKLEY, 1997; MANLY, 1997). O BST Tradicional (*Ordinary Bootstrap*) consiste em se “embaralhar” B vezes com reposição, a amostra existente e, então, inferir sobre as EI (doravante denominada estatística ou estimador). Em geral, utiliza-se B igual a 1000 para inferências, e nível de significância de 5%. Na amostra original determina-se a estatística ( $\hat{\theta}$  - Estatística de Interesse observada da amostra original) que será empregada para a análise do problema, e, a partir de geração das amostras BST, a estatística média é determinada ( $\hat{\theta}_B$  - Média das Estatísticas de Interesse a partir das simulações BST). Para pequenas amostras, de até 10 elementos, a determinação de todas as amostras possíveis BST pode ser implementada, sem a necessidade de simulação (FISHER e HALL, 1991; DAVISON e HINKLEY, 1997). Neste caso, a

abordagem BST balanceada (BSTbal) desponta como uma melhor alternativa do que o BST Tradicional, uma vez que possibilita que os elementos amostrais estejam presentes na mesma proporção (DAVINSON E HINKLEY, 1997; MAINLY, 1997).

A abordagem *Bootstrap* (incluindo suas variações) pode ser utilizada para determinação de limites de confiança, erros-padrão e testes de hipótese (ALMEIDA *et al.*, 2001; FISHER e HALL, 1991; DAVISON e HINKLEY, 1997; MANLY, 1997), de forma paramétrica ou não-paramétrica. A abordagem paramétrica é aquela em que a fdp da população é supostamente conhecida, e a simulação é feita de forma a seguir esta fdp. Caso a pressuposição de determinada fdp seja equivocada, os resultados podem ser tendenciosos. Na abordagem não-paramétrica, a simulação é feita a partir da geração de números aleatórios independentes e identicamente distribuída (iid), com fdp uniforme.

#### **II.4 – Revisão Bibliográfica sobre Utilização de RNA em Transplantes**

Na utilização de Redes Neurais Artificiais no contexto de Transplantes, DORSEY *et al.* (1997) desenvolveram uma RNA com algoritmo BP e compararam seus resultados com um modelo de Regressão Logística (RL) em transplantes de pâncreas (TXP), para determinar a probabilidade da perda do enxerto. A amostra constou de 117 pacientes, com TX simultâneo pâncreas-rim (SPR), 35 pacientes com TX pâncreas depois de rim (PDR) e 8 pacientes com TX pancreático somente (P). Destes, 23 pacientes tiveram perda do enxerto em um intervalo de até 3 meses pós-TX. Os pacientes foram operados no período entre junho de 1991 e fevereiro de 1996.

A Rede Neural Artificial foi desenvolvida utilizando *NEUROHELL2* (Ward Systems Group) e a rede escolhida como de desempenho ótimo possuía função de ativação Gaussiana nos neurônios ocultos e logística no único neurônio de saída. Doze variáveis de entrada foram selecionadas por recomendação de cirurgiões especialistas e, transplantes ou da literatura, e dicotomizadas (Tabela II.1). O desempenho dos modelos foi avaliado pela área sob a curva ROC (*Receiver Operating Characteristic*), teste  $\chi^2$  de Pearson e teste de Hosmer-Lemeshow. A influência de cada variável preditora também foi avaliada. O modelo de RL utilizou ponto de corte de 0,5 para a definição da sobrevida (<0,5) e perda do enxerto (>0,5). Para a RNA, foram dois pontos, para perda do enxerto (>0,70) e para sobrevida (<0,30), evitando-se os resultados intermediários. Utilizando as mesmas variáveis predictoras, o modelo RL obteve sensibilidade de 35,71% e especificidade de 100%,

enquanto a RNA, com arquitetura 12x20x1, obteve sensibilidade de 68% e especificidade de 96%. Os coeficientes de determinação foram 0,24 e 0,71 para a RL e a RNA, respectivamente.

**Tabela II.1** - Variáveis de entrada dicotômicas (*extraído de DORSEY et al., 1997*).

---

**Parâmetros**

---

*Tipo do transplante (PDR ou P)*

*Idade do doador > 35 anos*

*Idade do receptor > 50 anos*

*Sexo do Receptor*

*Sexo do Doador*

*Diferença do peso do receptor para o peso do doador > 50 lb*

*Transfusões Sangüíneas*

*Tabagismo*

*Alcoolismo*

*Centro de procura de doadores não-local*

*Histórico de diabetes por mais de 25 anos*

*Compatibilidade HLA (Human Leukocyte Antigens) com 2 Loci coincidentes*

*Diálise prévia*

---

SHEPPARD *et al.* (1999) investigaram várias topologias de PMC (com uma camada oculta e com algoritmo BP) para previsão de infecção de Citomegalovirus (CMV) após TXR em 548 pacientes (77 desenvolveram Citomegalovirose pós-TX), operados entre 01/01/1986 e 1/12/1994 no Reino Unido. Dezesete variáveis (ou expandidas para 42, onde as variáveis categóricas foram transformadas em variáveis binárias) foram selecionadas por especialistas da área (Tabela II.2). Os dados dicotômicos (sexo, status em relação à infecção por CMV, tipo de doador, receptor diabético, profilaxia contra CMV, infecção por CMV pós-TX, tipo do transplante) foram codificados como 0 ou 1, enquanto os demais (idades, HLA Compatibilidade, PRA, esquema imunossupressor) foram categorizados. O status da relação entre doador (D) e receptor (R) em relação à infecção prévia por CMV eram D+/R+ = 163, D+/R- = 130, D-/R+ = 132 e D-/R- = 123, onde o símbolo “+” significa infectado. A função de ativação utilizada foi a sigmóide, as taxas de aprendizagem foram 0,3 e 0,15 para as

camadas oculta e de saída, respectivamente. A constante de momento foi ajustada para 0,4.

**Tabela II.2** - Variáveis de entrada (*extraído de SHEPPARD et al., 1999*).

---

**Parâmetros**

---

*Idade do Receptor*

*Idade do Doador*

*Sexo do Receptor*

*Sexo do Doador*

*Status do Receptor em relação à infecção prévia por CMV*

*Status do Doador com relação à infecção prévia por CMV*

*Nº de episódios de rejeição*

*Nº de transplantes anteriores*

*Tipo do Doador (cadavérico ou vivo relacionado)*

*Compatibilidade HLA*

*Tipo do esquema imunossupressor*

*Painel de Reatividade (PRA) Pré-TX mais recente*

*PRA Pré-TX mais elevado*

*Receptor com Diabetes*

*Profilaxia contra CMV*

*Infecção por CMV detectada pós-TX*

*Tipo do TX (rim ou rim/pâncreas)*

---

As redes testadas por SHEPPARD *et al.* (1999) possuíam 3, 5, 7, 10, 15 e 20 neurônios na camada oculta, com 1 ou 2 neurônios na camada de saída. Com a utilização de um neurônio na camada de saída a rede mostrou-se excessivamente lenta, razão pela qual foram escolhidos dois neurônios na camada de saída. O “software” utilizado foi o *NEURALWORKS PROFESSIONAL II PLUS* (Neuralware Inc.), sendo que a análise do desempenho do modelo foi feita utilizando-se a área sob a curva ROC. As saídas variavam continuamente entre 0 e 1, de forma a classificar a probabilidade de ocorrência da Citomegalovirose (que consistia na infecção por CMV associada aos sintomas característicos). Como existiam muito mais casos de pacientes que não desenvolveram Citomegalovirose, os autores estabeleceram

algumas estratégias para treinar e testar a rede tentando “balancear” os dados entre os dois grupos:

**método 1):** foram selecionados 39 pacientes aleatoriamente nos dois grupos, com 10 diferentes conjuntos de teste e treinamento, associados a 10 redes diferentes (1 para cada conjunto de treinamento), com topologias 17x3x2, 17x5x2, 17x7x2, 17x10x2. As redes foram treinadas com 1 conjunto e testadas com todos os 10 conjuntos de teste. A saída foi determinada de acordo com a expressão  $R = 0,5 - (saída2 - saída1) \div 2$ . Se o valor de R tendesse a 1, a probabilidade de ocorrência da doença seria elevado.

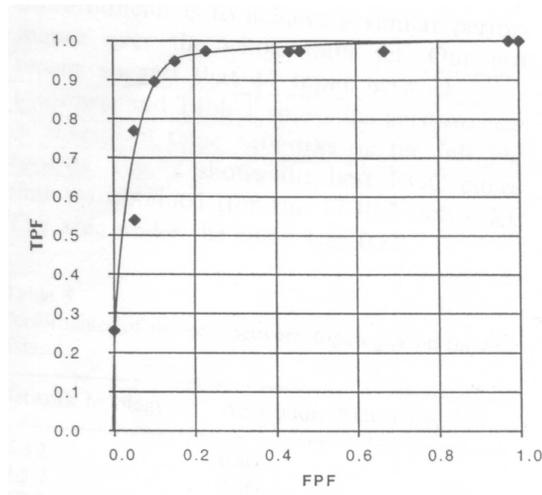
**método 2):** foi utilizado o Método-Deixe-Um-De-Fora para todo o conjunto, formando-se então 548 conjuntos de treinamento e 548 conjuntos de teste, com topologias de 5, 10, 15, 20 25 e 30 neurônios na camada oculta. Como o tempo de treinamento e teste neste processo é muito grande, e ainda não sendo possível balancear o número de casos de ocorrência da doença com os casos não afetados, dois métodos foram adicionalmente avaliados utilizando esta abordagem.

- **(2a):** foram obtidos 154 conjuntos de treinamento e teste, utilizando os 77 pacientes que desenvolveram a doença e outros 77 que não a desenvolveram, selecionados aleatoriamente;

- **(2b):** foram gerados 933 conjuntos de treinamento e teste, utilizando 6 cópias combinadas dos pacientes que desenvolveram Citomegalovirose, totalizando  $6 \times 77 = 462$ , somados aos 471 que não desenvolveram a doença, “balanceando” satisfatoriamente os dois grupos, e utilizando todos os dados.

Neste estudo de SHEPPARD *et al.* (1999), a análise do desempenho baseou-se na área sob a curva ROC, que possibilita avaliar a sensibilidade (quantidade de pacientes que desenvolveram a doença e que foram corretamente classificados como tal) em relação ao resultado Falso-Positivo (aqueles pacientes que não desenvolveram a doença, mas foram incorretamente classificados como doentes). Cada par Sensibilidade (TPF, *True-Positive Fraction*) x Falso-Positivo (FPF, *False-Positive Fraction*), de acordo com limiares estabelecidos para a saída da rede, foi determinado e caracterizou um ponto da curva ROC, com coordenadas FPF x TPF. Pelo método 1, usando o vetor de entradas com 17 variáveis, o melhor desempenho foi obtido com a rede 17x10x2, com área sob ROC de 0,93 (Figura II.1). Utilizando 42 variáveis de entrada, o melhor desempenho foi obtido com topologia 42x20x2, área sob ROC de

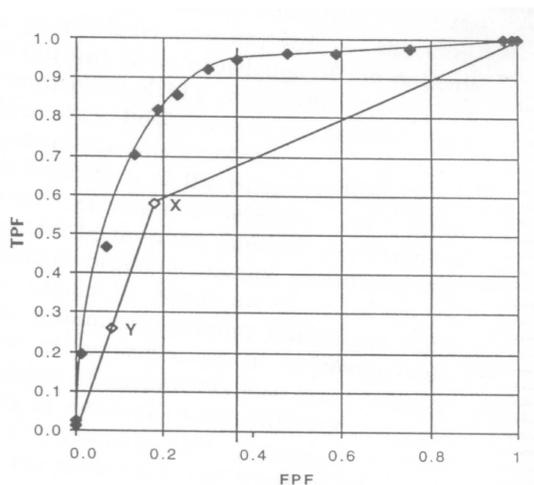
0,88 (método 2b). Os autores atribuíram o melhor desempenho como sendo da rede 17x10x2 (método 2b), uma vez que a mesma rede sob o método 1 não contemplava toda a base de dados. Neste caso, a área sob ROC foi de 0,88.



**Figura II.1** – Curva ROC, para a rede 17x10x2, com limiares de aceitação da saída variando de 0, 0,01, 0,02, 0,1, 0,15, 0,2, ..., 0,9, 0,99 e 1, treinamento segundo método 1 (extraído de SHEPPARD *et al.*, 1999).

SHEPPARD *et al.* (1999) também avaliaram o desempenho clínico com base na curva ROC. O número de pacientes com maior possibilidade de desenvolver a doença (D+/R-) é de 130 pacientes. Destes, 45 desenvolveram Citomegalovirose, caracterizando  $TPF = 45/77 = 0,58$  e  $FPF = 85/471 = 0,18$ , sendo o ponto X da figura II.2. Os autores consideraram que o efeito da profilaxia teria uma tendência de deslocar este ponto, mas não muito, devido ao fato de que a profilaxia não foi muito efetiva (86 receberam profilaxia e 34 desenvolveram a doença, e dos 44 que não receberam, só 11 desenvolveram Citomegalovirose). Do grupo que não desenvolveu Citomegalovirose (471), 52 pacientes receberam profilaxia, enquanto que no outro grupo (77), 34 pacientes foram contemplados com a droga profilática. Dos 419 pacientes que não desenvolveram e também não receberam profilaxia, 33 pertenciam ao grupo D+/R-, ou seja,  $FPF = 0,08$ . Dos 43 que desenvolveram e não receberam profilaxia, 11 também faziam parte do mesmo grupo, então  $TPF = 0,26$ . Este é o ponto Y da Figura II.2, sendo a origem caracterizada como a suposição de que não haverá

citomegalovirose em nenhum paciente e o ponto (1,1) com sendo o contrário. Esta curva, embora possuindo aproximações, caracterizaria uma envoltória aproximada da avaliação clínica, ficando, portanto, com um desempenho, segundo a área sob a ROC, inferior ao da RNA. Com vistas a melhorar a modelagem das redes, em relação às variáveis de entrada, os autores utilizaram programa intitulado *PREDICT*, que otimiza a arquitetura da rede em relação às variáveis de entrada, utilizando Algoritmos Genéticos (AG). A remoção de 7 variáveis de entrada (n° de transplantes anteriores, tipo do doador, tipo de transplante, PRA mais recente, sexo do receptor, receptor com diabetes e profilaxia contra CMV) propiciou uma área sob ROC de 0,963, com sensibilidade de 0,9 e especificidade de 0,83, com menos variáveis de entrada do que o estabelecido previamente.



**Figura II.2** – Curva ROC, para a rede 17x10x2, com limiares de aceitação da saída variando de 0, 0,01, 0,02, 0,1, 0,15, 0,2, ..., 0,9, 0,99 e 1, considerando toda a base de dados e treinamento segundo método 2. Pontos X e Y correspondem à avaliação clínica (extraído de SHEPPARD *et al.*, 1999).

FURNESS *et al.* (1999a e 1999b), projetaram uma rede PUC (Matlab Neural Networks Toolbox) com 12 variáveis de entrada, mostradas na Tabela II.3, a partir de aspectos histológicos biopsiados de enxertos renais. Se a saída do combinador linear fosse maior do que um limiar estabelecido, era feita a classificação como rejeição ou não. Com treinamento de 125 casos, e teste em 21 biópsias, a rede diagnosticou corretamente 19 casos de rejeição aguda. A Regressão Logística, realizada para

comparação, diagnosticou 8 casos corretamente. Os grupos de treinamento e teste foram assim separados: 100 casos (43 de rejeição) foram anteriormente diagnosticados facilmente, 25 casos (19 de rejeição) diagnosticados como “borderline”, e 21 casos (10 de rejeição) de difícil diagnóstico, todos considerados à época das biópsias. O protocolo consistiu do treinamento inicial dos 100 casos, com teste da rede dos demais.

**Tabela II.3** - Aspectos Histológicos das biópsias dos enxertos renais (*modificado de FURNESS et al., 1999*).

<b>Aspectos Histológicos</b>	<b>Escore</b>
<i>Tubulite</i>	0-3, de acordo com <i>Banff</i>
<i>Arterite Intimal</i>	0-3, de acordo com <i>Banff</i>
<i>Infiltrados Intersticiais</i>	% da área da biópsia com infiltração de linfócitos
<i>Edema</i>	% da área mostrando edema intersticial
<i>Hemorragia Intersticial</i>	0-3, graduação específica
<i>Glomerulite Aguda</i>	0-3, de acordo com <i>Banff</i>
<i>Linfócitos Ativados</i>	0-3, graduação específica
<i>Venulite</i>	0-3, graduação específica
<i>Eosinófilos</i>	Nº por campo mais infiltrado
<i>Células Plasmáticas</i>	Nº por campo mais infiltrado
<i>Aderência de células mononucleares endoteliais arteriais</i>	Presença ou Ausência

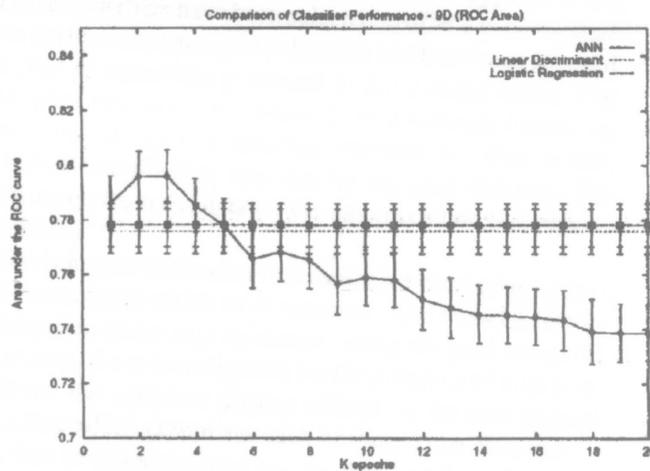
*Banff* é a classificação utilizada para lesões histológicas observadas nos enxertos renais

Usando o conjunto de 25 casos “borderline” para teste, FURNESS *et al.* (1999a) obtiveram desempenho desanimador, pois somente 4 casos foram corretamente diagnosticados, enquanto que ao utilizar o conjunto de 21 casos *difíceis*, 11 casos foram corretamente diagnosticados. Na avaliação dos autores, tal desempenho deveu-se ao fato de que o conjunto de treinamento era composto de casos óbvios de rejeição. Completando-se o conjunto de treinamento com os 25 casos “borderline”, o desempenho da rede alterou-se substancialmente, detectando corretamente 19 casos (de 21). Deste modo, eles concluíram que conjuntos de treinamento com casos simples não terão bom desempenho em casos mais complexos, devendo estas

informações estarem presentes também no treinamento. Acrescentam ainda que, devido à natural capacidade da RNA de integrar diferentes tipos de entrada, outros tipos de variáveis poderiam ser incorporados ao modelo, tais como tempo pós-TX, taxa de variação de creatinina ou níveis séricos de drogas imunossupressoras.

No outro artigo, FURNESS *et al.* (1999b) também enfatizaram aplicabilidade da RNA, comparando com o desempenho de 31 patologistas especializados em biópsias de enxertos renais. O melhor desempenho dentre os patologistas alcançou um número de acertos de 18 casos de rejeição (em um total de 21).

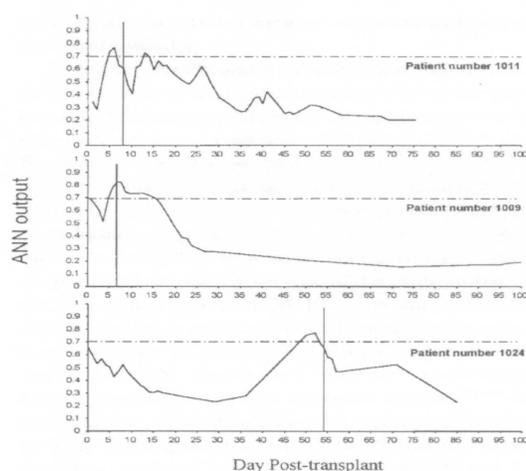
A aplicação de RNA em transplantados hepáticos (TXH) foi estudada por MELVIN *et al.* (2000) e Hughes *et al.* (2001), sendo estes trabalhos de um mesmo grupo de pesquisadores. MELVIN *et al.*, 2000 utilizaram PMC, Análise de Discriminantes Lineares (ADL) e RL para monitorização de rejeição aguda (80 pacientes e 15 retransplantes). Para monitorar a evolução pós-transplante foram empregadas as variáveis alfa-Glutathione S-Transferase (GST), Fosfatase Alcalina (FA), Alanina Transaminase (ALT), Bilirrubinas (BILI), seus respectivos gradientes (gr's) e dia pós-TX (significando aqui, simplesmente, a taxa de variação). Três tipos de vetores de entrada foram utilizados, de 4 variáveis (GST, FA, ALT e BILI), de 8 variáveis (GST, FA, ALT, BILI, grGST, grFA, grALT e grBILI) e de 9 variáveis (o mesmo que o anterior acrescido do dia pós-TX). Os gradientes foram calculados de acordo com um filtro média móvel (MA, *Moving Average*) de 3ª ordem, com os pontos ajustados posteriormente por uma *spline* cúbica, e finalmente, calculada a primeira derivada do polinômio resultante para o intervalo de tempo entre as amostras. Como a função densidade de probabilidade das variáveis séricas é log-normal, os dados foram transformados logaritmicamente. Do total de pacientes, 30% foram selecionados aleatoriamente para formar o conjunto de validação. Cem conjuntos de treinamento e validação pseudo-aleatórios foram criados para comparação entre os métodos. As arquiteturas dos PMC estabelecidas para os vetores de entrada de 4, 8 e 9 variáveis foram 4x5x1, 8x10x1, 9x11x1, respectivamente. As funções de ativação das camadas ocultas e de saída são funções Logísticas e a rede é totalmente conectada. O algoritmo de treinamento foi o BP, com decaimento de pesos, e critério de parada antecipada baseado na área sob a curva ROC. A Taxa de Aprendizagem e Constante de Momento finais foram 0,2 e 0,5, respectivamente. Dentre as arquiteturas, o desempenho do PMC demonstrou ser o melhor, de acordo com a área sob a curva ROC (Figura II.3) e do coeficiente de determinação, cujas faixas foram 0,74-0,78 e 0,76-0,80 para RL e RNA, respectivamente



**Figura II.3** - Área sob a curva ROC em função de n° de épocas. A diminuição do desempenho é creditada pelos autores à sobreparametrização (extraído de MELVIN *et al.*, 2000).

Posteriormente, HUGHES *et al.* (2001) usaram uma RNA com 5 entradas (ALT, grALT, BILI, grBILI e dias pós-TX) e uma camada oculta (n° de neurônios não especificado) para predição de rejeição aguda em TXH (116 pacientes, 124 enxertos). Um único neurônio de saída, com valores de saída entre 0 e 1, representou a probabilidade de rejeição. Tal rede já havia sido treinada e testada anteriormente em outro grupo de transplantados hepáticos, conforme referência citada no estudo. Os procedimentos cirúrgicos foram realizados entre julho de 1995 e outubro de 1997, e as medições dos marcadores de função hepática foram medidos diariamente, durante 3 meses. Setenta episódios de rejeição aguda (em 59 enxertos) foram diagnosticados. A análise de sensibilidade e especificidade foi feita utilizando-se a área sob a curva ROC, tanto para a RNA quanto para os valores de ALT, BILI e dia pós-TX. Após várias configurações, a RNA foi significativamente mais específica no diagnóstico de rejeição aguda do que ALT, BILI e dia pós-TX ( $\chi^2$ ,  $p < 0,0001$ ). A análise da sensibilidade não revelou diferenças significativas entre as variáveis, exceto para dia pós-TX ( $\chi^2$ ,  $p < 0,0001$ ). A RNA foi capaz de diagnosticar a ocorrência de rejeição mais cedo do que o diagnóstico por biópsia, através do seu pico de valores de saída, representando o risco mais alto de rejeição aguda (mediana de 1 dia, 95% de intervalo de confiança, vide Figura II.4). Este valor foi o utilizado para análise da ROC. O limiar ótimo da saída

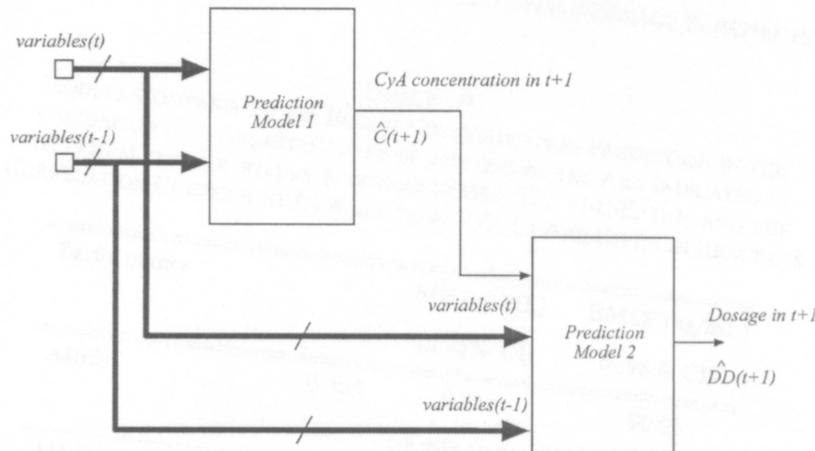
da rede foi determinado como sendo aquele que originou a maior soma entre sensibilidade e especificidade.



**Figura II.4** – Gráfico de 3 pacientes que experimentaram episódio de rejeição aguda. A linha cheia vertical foi o diagnóstico médico. A linha tracejada foi o limiar ótimo da rede para o paciente (*extraído de HUGHES et al., 2001*).

Redes Neurais Artificiais dinâmicas na predição de dosagem de uma droga imunossupressora chamada Ciclosporina A (CsA) em Transplante Renal foram usadas por CAMPS-VALLS *et al.* (2003). O modelo consistiu de duas RNA interligadas. Uma rede foi utilizada para modelar o comportamento do paciente frente à dosagem da CsA, cujas entradas foram concentração sérica de CsA, dose terapêutica de CsA, creatinina sérica, idade, sexo, peso corporal total e dia pós-TX. A saída desta rede estimava a concentração de CsA para ajuste da próxima dose, e pertencia, junto com as demais variáveis de entrada, ao vetor de entrada da rede que estimava a próxima dose a ser ministrada ao paciente (vide Figura II.5). Trinta e dois pacientes serviram como amostras (22 como treinamento – 364 medições, e 10 como validação – 217 medições). O esquema imunossupressor era composto de CsA, Prednisona (Pred) e Micofenolato Mofetil (MMF). Os tipos de RNA utilizadas foram o PMC (acrescido de um vetor de elementos de atraso unitários), uma rede FIR (RNFIR) e uma rede recorrente de Elman (EL), todas desenvolvidas em ambiente MATLAB®. Um modelo ARMA foi também desenvolvido para análise comparativa. As concentrações séricas de CsA foram medidas em diferentes ocasiões, onde eram também obtidos os demais

dados, configurando as séries temporais de entrada. A combinação dos modelos foi utilizada para aumentar a eficácia.



**Figura II.5** - Esquema das RNA utilizadas (extraído de Camps-Vall et al., 2003).

CAMPS-VALLS *et al.* (2003) utilizaram modelos ARMA baseados em critérios de informação (Akaike, Akaike Bayesiano, dentre outros) e em autocorrelação simples e parcial para a predição de níveis séricos de concentração e dosagem terapêutica foram AR(3) e AR(1), respectivamente. O PMC foi iniciado com pesos e Taxa de Aprendizagem variando entre 0,001 e 0,3, o n° máximo de neurônios na camada oculta foi de 20. Para a RNFIR, o n° de neurônios ocultos, o n° de derivações por conexão sináptica e a Taxa de Aprendizagem variaram de 2 a 25, 1 a 4, 0,0001 a 0,01, respectivamente. Já a rede EL utilizou Constante de Momento  $> 0,8$ . Não foram citadas as funções de ativação e nem o n° de neurônios na camada de saída. O critério de avaliação utilizado foi o do RMSE (*Root Mean Square Error*) e o Erro Médio (ME, *Mean Error*) de predição para verificação do *Bias* (tendenciosidade). A eficácia dos modelos foi avaliada por ANOVA (*Analysis of Variance*). Foi feita a introdução de ruído Gaussiano com média zero e diferentes variâncias para avaliar a robustez dos modelos. Os resultados das redes sem a introdução do ruído não demonstraram diferença significativa, comparando a média dos erros de predição ( $p=0,091$ ) ou a média dos erros de predição absolutos ( $p=0,999$ ). Já com a introdução do ruído, a RNFIR e EL tiveram melhor desempenho. No esquema interligado, o melhor

desempenho foi dado pela utilização da RNFIR para a predição da concentração de CsA e da EL para a dosagem terapêutica.

Utilizando transformações das variáveis de entrada, Linearização e Agrupamento mais Quantização, DVORCHIK *et al.* (1996) compararam os desempenhos de duas Redes Neurais Artificiais Multi-Camadas aplicadas a 155 pacientes de Transplante Hepático. A ênfase desse trabalho foi na comparação de diferentes métodos de tratamento de dados perdidos. Já SHOSKES *et al.* (1998) utilizaram RNA para predição de rápido funcionamento do enxerto de doador cadavérico e predição de rejeição aguda em Transplante Renal. Eles testaram uma Rede Neural Artificial treinada com algoritmo de retropropagação do erro configurada pelo programa *NEURALYST* (Cheshire Engineering), treinada em 100 pacientes e testada em 20 pacientes que receberam transplantes cadavéricos. As variáveis de entrada referentes ao doador foram idade, *causa mortis*, história de hipertensão, uso de inotropa, débito urinário, creatinina sérica inicial e final, variação de creatinina. As variáveis do receptor foram tempo de isquemia do enxerto, número prévio de transplantes, sensibilização por anticorpos e Compatibilidade HLA. A saída foi considerada como sendo funcionamento tardio do enxerto (DGF, *Delayed Graft Function*) e creatinina sérica menor do que  $2,5 \text{ mg.dL}^{-1}$ , após 10 dias. O programa determinou, através de algoritmo de otimização genético, duas camadas ocultas, com 20 e 10 neurônios, propiciando aprendizagem de 100% dos exemplos de treinamento, alcançando um valor mse de 0,0036. Na amostra de teste, a rede conseguiu predizer 80% da saída DGF e 100% da saída creatinina  $< 2,5 \text{ mg.dL}^{-1}$ , após 10 dias.

Também em Transplante Renal, SIMIC-OGRIZOVIC *et al.* (1999) determinaram, dentre 33 variáveis previamente selecionadas, três grupos de variáveis preditivas, de acordo com seu grau de influência (forte, fraca e sem influência) na predição de rejeição crônica em 27 pacientes transplantados renais. HSIANG-YIN *et al.* (1999) utilizaram RNA dinâmica com algoritmo BP para predição de dosagem de Tacrolimus (um medicamento imunossupressor) durante os primeiros seis meses em TXH, em 32 pacientes. A seleção das variáveis de entrada da rede foi feita utilizando-se AG. O desempenho da RNA foi considerado superior ao da Regressão Linear Múltipla, que foi utilizada para comparação. As concentrações de Tacrolimus no conjunto de teste tiveram proporção de acertos de 84% e 66% para a Rede Neural Artificial e a Regressão Linear Múltipla, respectivamente.

Esta revisão mostra que existem poucos trabalhos publicados na área de Transplantes, apesar do avanço da utilização das Redes Neurais Artificiais na área

médica (HUGHES *et al.*, 2001). Entretanto, os analisados aqui apontam para a aplicação crescente das Redes Neurais Artificiais na solução de distintos problemas da área de transplantes e pretendemos, com o presente trabalho, introduzir novas contribuições.

## CAPÍTULO III

### Análise Exploratória de Dados de Transplantados Renais

#### III.1 - Introdução

Em geral, as variáveis selecionadas como preditoras de um modelo são baseadas na experiência do pesquisador. Se a base de dados na qual se encontram as variáveis preditoras é retrospectiva, sempre é possível deparar-se com a ocorrência de dados perdidos para as variáveis que se deseja utilizar. A estratégia usual para este problema é o descarte dos casos com dados incompletos, porém, dependendo da quantidade de casos remanescentes, o modelo preditivo pode apresentar menor acurácia e resultados tendenciosos (OOSTENBRINK *et al.*, 2003).

O conjunto amostral deste trabalho abrangeu o período de janeiro de 1986 até dezembro de 2000, totalizando 556 pacientes transplantados renais (receptores), selecionados do Hospital Universitário Pedro Ernesto, da Universidade Estadual do Rio de Janeiro. Dados anteriores a 1986 não foram aproveitados devido à introdução da Ciclosporina A (CsA) nesse ano, um ponto importante na sobrevida dos enxertos renais (MATSUI *et al.*, 2001; CAMPS-VALLS *et al.*, 2003; ÁVILA *et al.*, 2004). Na análise exploratória dos dados foram utilizados histogramas e diagramas *boxplot*. Para inferir sobre a Gaussianidade utilizou-se histogramas, diagramas qq-plot e o teste de Shapiro-Wilk (JOHNSON e WICHERN, 1988; LOONEY, 1995). O nível de significância para todos os testes estatísticos empregados foi  $\alpha = 5\%$ .

#### III.2 – Definição das variáveis dependente e preditoras

A variável dependente foi definida como a *Sobrevida do Enxerto* (SE) em meses. Mesmo quando não houve acompanhamento do paciente com enxerto funcionando a partir de determinada data (dado censurado), a sobrevida foi considerada como sendo a diferença entre a última data de acompanhamento e a data do transplante.

A data limite de realização do transplante foi estabelecida como sendo 01/12/1995 (neste ano, somente dois receptores tiveram as variáveis preditoras disponíveis completas, porém sem informações sobre a sobrevida do enxerto e, portanto, foram descartados. Deste modo, somente foram considerados receptores até dezembro de 1994). Na organização dos dados, as colunas representavam as

informações obtidas dos receptores. Oitenta e duas variáveis foram coletadas à época da cirurgia, incluindo nomes e iniciais dos receptores, telefones de contato etc. A fim de garantir a privacidade dos receptores, os mesmos foram codificados numericamente (vide Apêndice B).

Com o critério adotado, 297 receptores com doadores vivos foram incluídos para este trabalho. Deste total, algumas variáveis tiveram que ser descartadas, face à alta proporção de dados incompletos. Por exemplo, variáveis como sorologia para Antígeno da Hepatite B (HbsAG) e vírus da Hepatite C (HCV), importantes causas de morbimortalidade em Transplantados Renais (TXR) (SALOMÃO-FILHO *et al.*, 1996) não entraram no estudo devido à proporção de dados incompletos encontrados no período (67 e 77%, respectivamente). A sorologia para Citomegalovírus (CMV), outra importante variável preditora (SHEPPARD *et al.*, 1999), também teve alta proporção de dados incompletos, bem como variáveis indicativas de rejeição (>50%), todas portanto descartadas. Somente 109 receptores com dados completos foram selecionados (ENNET *et al.*, 2001).

As variáveis *Human Leukocyte Antigens* (HLA) Loci A e B foram transformadas em uma única variável categórica (*Compatibilidade HLA*). Se os dois Loci para HLA A e B eram iguais, *Compatibilidade HLA* foi codificado como 3 (doador idêntico), se apenas um Locus era igual, código 2 (doador haplo), e se não havia Loci coincidentes, código 1 (doador não relacionado). As variáveis *Pulso Corticóide*, *OKT3* e *Anticorpo Anticlinal* foram transformadas na variável dicotômica *Rejeição*, uma vez que a utilização de quaisquer uma destas implica em tratamento desta complicação.

Todas as categorias das variáveis dicotômicas consideradas como a favor da sobrevida do enxerto tiveram seus valores definidos como 1. A imunossupressão de manutenção foi dicotomizada como 1 (se o paciente utiliza Ciclosporina A) ou 0 (sem Ciclosporina A). Todos os receptores utilizaram Prednisona, e alguns receptores utilizaram Ciclosporina A em esquema duplo ou tríplice (MATSUI *et al.*, 2001; CAMPS-VALLS *et al.*, 2003). Devido à menor massa renal do sexo feminino, o receptor feminino foi classificado com valor 1 e o doador feminino com valor 0 (SOLÁ *et al.*, 2002; GOLDFARB-RUMYANTZEV *et al.*, 2003). Assim, no caso de doador masculino com receptor feminino, ambas variáveis assumem valor 1. *Raça* não-afrodescendente assumiu valor 1, tanto para doador quanto para receptor (GOLDFARB-RUMYANTZEV *et al.*, 2003; Tabela 5.8b, ano 2003, da *Organ Procurement and Transplantation Network*, publicada pela *United Network for Organ Sharing* – UNOS). Tipo de diálise CAPD (*Continuous Ambulatory Peritoneal Dialysis*) teve valor 1 (GOLDFARB-

RUMYANTZEV *et al.*, 2003). As variáveis que foram inicialmente consideradas para o trabalho estão listadas na Tabela III.1. A relação dos 109 receptores considerados neste estudo é mostrada no Apêndice B.

**Tabela III.1- Variáveis preditoras preliminarmente definidas**

<b>Variáveis</b>	<b>Símbolo</b>	<b>Tipo da Variável</b>
<i>Idade Receptor</i> (anos)	E1	Contínua
<i>Idade Doador</i> (anos)	E2	Contínua
<i>Sexo Receptor</i>	E3	Dicotômica, f = 1
<i>Sexo Doador</i>	E4	Dicotômica, f = 0
<i>Raça Receptor</i>	E5	Dicotômica, afro-descen = 0
<i>Raça Doador</i>	E6	Dicotômica, afro-descen = 0
<i>Transplantes Prévios</i>	E7	Numérica
<i>Transfusões Sangüíneas Receptor</i>	E8	Numérica
<i>Duração da Diálise</i> (meses)	E9	Numérica
<i>Tipo da Diálise</i> (HD, CAPD)	E10	Dicotômica, HD = 0
<i>Rejeição</i>	E11	Dicotômica, não = 1
<i>Compatibilidade HLA</i> (Doador Vivo)	E12	Categ, Id = 3, Haplo = 2, Não-Relacionado = 1
<i>Esquema Imunossupressor</i>	E13	Dicotômica, uso de CSA = 1

### III.2.1 – Variáveis contínuas

A *Idade dos Receptores* (E1) variou entre 6 e 66 anos, sendo que 93 destes (~70%) encontram-se na faixa etária de 20 a 57 anos. O histograma e o *qq-plot* da Figura III.1 (Linha 1) sugerem a Gaussianidade desta variável ( $\alpha = 5\%$ , teste Shapiro-Wilk com  $W = 0,99$  e  $p = 0,61$ , indicando a não rejeição da hipótese nula. Embora dois doadores (64 e 66 anos, respectivamente) pudessem ser considerados *outliers*, o diagrama *boxplot*, para o qual 95º percentil é 55,6 anos, não confirma esta indicação.

A variável *Idade do Doador* (E2) possuía faixa de variação de 19 a 67 anos e 95º percentil em 56,6 anos. Na Figura III.1 (Linha 2) são mostrados o histograma, *boxplot*, *qq-plot* e o diagrama de espalhamento desta variável. Por Shapiro-Wilk não houve indicação de Gaussianidade ( $W = 0,97$ ,  $p = 0,01$ ) e nem o *boxplot* indica probabilidade de *outliers*.

Por outro lado, para a variável *Transfusões Sangüíneas* (E8), o histograma sugere uma distribuição exponencial e o *boxplot* mostra que existe probabilidade de existência de *outliers*, dois com 26 transfusões (transplantados em 1986) e um terceiro com 24 (transplantado em 1987), sendo o 95º percentil igual a 20,6. Além destes, mais três receptores ultrapassaram o 95º percentil (transplantados em 1987, 1988 e 1990, respectivamente). O primeiro quartil e a mediana têm valores limítrofes de 2 e 5 transfusões, respectivamente, ou seja, quase metade dos receptores (51) tem até 5 transfusões sangüíneas exclusive (Figura III.1, Linha 3).

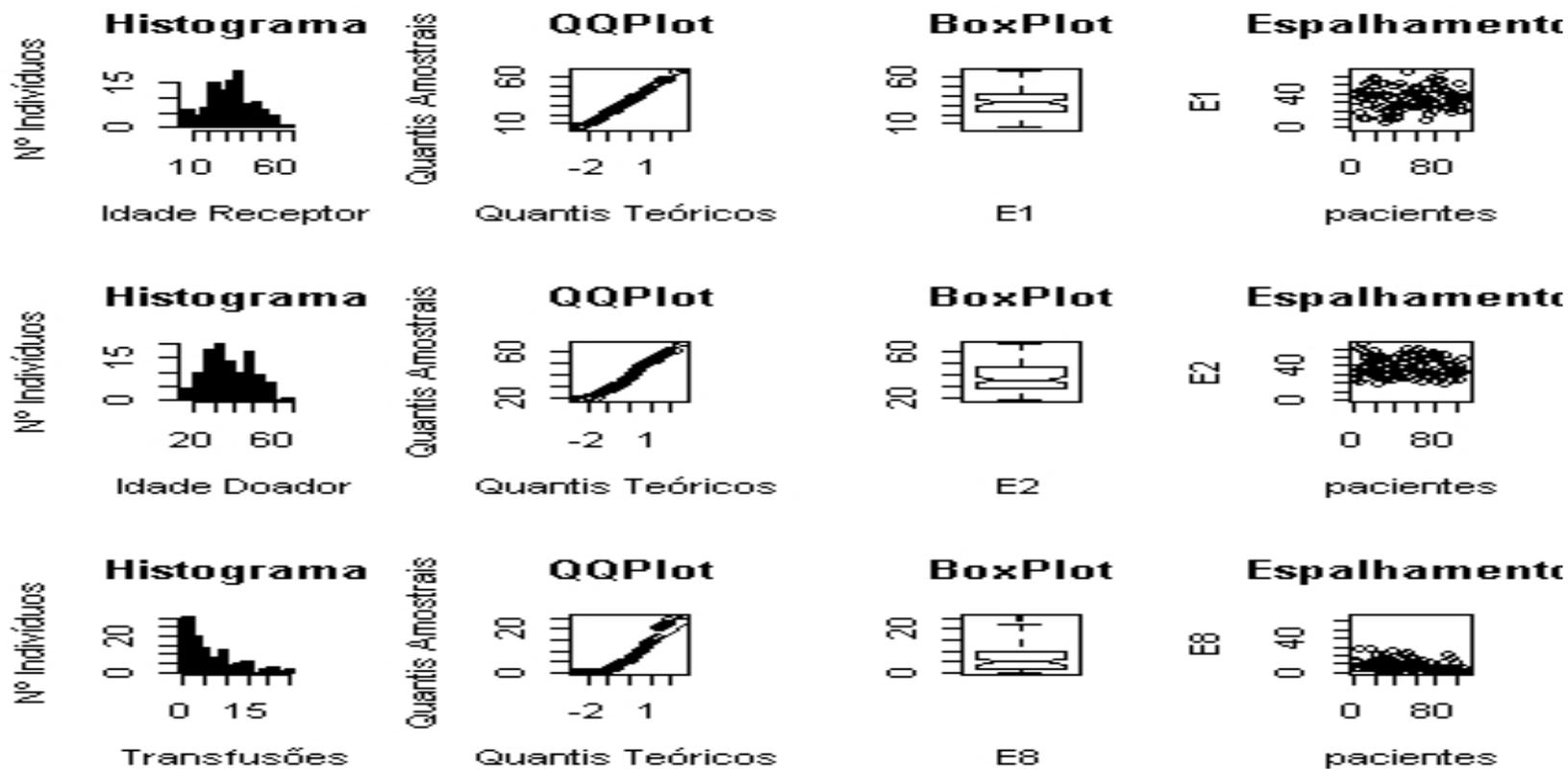
A variável *Duração da Diálise* (E9) possui características similares à variável E8, sugerindo portanto uma fdp exponencial (Figura III.2, Linha 1). Um receptor possui a maior duração de diálise antes do transplante (78 meses, transplantou em 1988), o que parece ser *outlier* para receptores com doadores vivos. O 95º percentil é 43,5, sendo 9,5 e 17,1 meses os valores para o primeiro e segundo quartis, respectivamente.

A maior parte dos valores da variável dependente *Sobrevida do Enxerto* (Figura III.2, Linha 2) eram inferiores a 109 meses. Embora o gráfico de espalhamento e o histograma sugiram a existência de *outlier* (1 paciente), o *boxplot* não ratifica esta hipótese. O 95º percentil é 109,2, com sobrevida máxima de 146,1 meses. Alguns receptores (69) são censurados, ou seja, são receptores que não possuem informações a partir de determinada data. Teste de Shapiro-Wilk não indicou Gaussianidade ( $W = 0,95$ ,  $p = 0,0005$ ). Como o período de acompanhamento mínimo foi de 5 anos (de 1995 a 2000), a variável *Sobrevida do Enxerto* foi limitada a 60 meses.

O 95º percentil das variáveis preditoras foi escolhido como “ponto de saturação” das variáveis contínuas visando a inclusão de possíveis *outliers*. Embora a variável E7 fosse contínua, somente 2 receptores tiveram transplantes anteriores. O resumo da análise das variáveis contínuas está mostrado na Tabela III.2.

**Tabela III.2 – Resumo das variáveis contínuas**

<b>Variável</b>	<b>Saturação</b>	<b>Média</b>	<b>Mediana</b>	<b>Faixa</b>
<i>Idade Receptor</i> (anos)	55,6	33,3	33	6-66
<i>Idade Doador</i> (anos)	56,6	38	36	19-67
<i>Transfusões Sangüíneas</i>	20,6	6,8	5	0-26
<i>Duração Diálise</i> (meses)	43,5	19,6	17,1	0,3-78,4
<i>Sobrevida do Enxerto</i> (meses)	60,0	52,2	48,7	0-146,1



**Figura III.1** – Características das variáveis predictoras. Linha 1: *Idade do Receptor*. Linha 2: *Idade do Doador*. Linha 3: *Transfusões Sangüíneas*

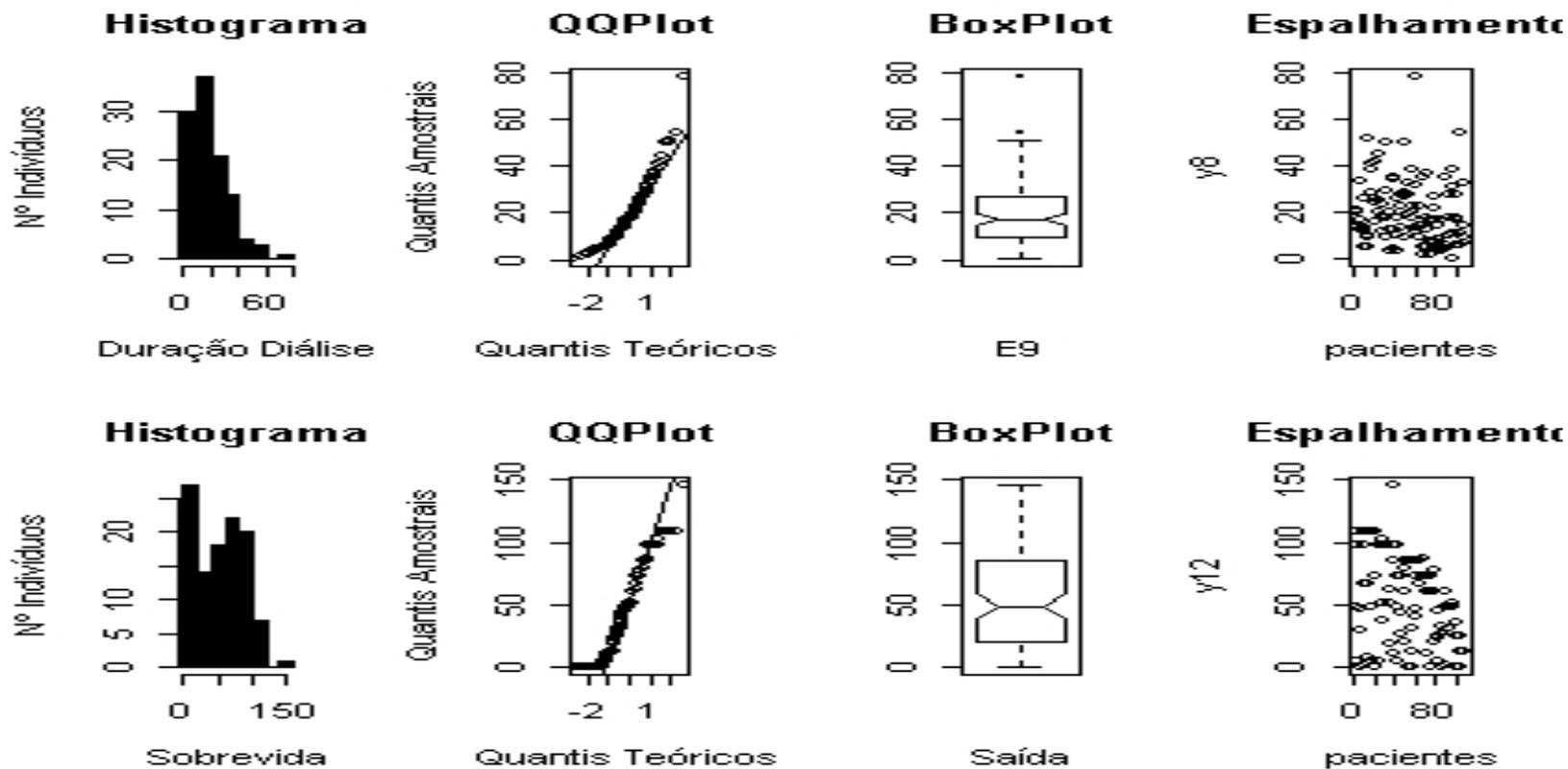


Figura III.2 – Características da variável preditora *Duração da Diálise* (Linha 1) e da variável dependente *Sobrevida do Enxerto* (Linha 2).

### III.2.2 – Variáveis dicotômicas

As variáveis dicotômicas E3, E4, E5, E6, E10, E11 e E13 (vide Tabela III.3) possuíam percentagens de valor 1 conforme mostrado. A variável E11 não possuía valor 0 em nenhum dos receptores selecionados.

**Tabela III.3** – Proporção do valores 1 das variáveis dicotômicas. Valor 1 contribui para o sucesso da *Sobrevida do Enxerto*

Variáveis	Siglas	Proporção
<i>Sexo Receptor</i>	E3, fem = 1	73/109
<i>Sexo Doador</i>	E4, masc = 1	65/109
<i>Raça Receptor</i>	E5, Não-Afrodesc = 1	50/109
<i>Raça Doador</i>	E6, Não-Afrodesc = 1	54/109
<i>Tipo da Diálise (HD, CAPD)</i>	E10, CAPD = 1	3/109
<i>Rejeição</i>	E11, não = 1	109/109
<i>Esquema Imunossupressor</i>	E13, uso de CSA = 1	76/109

### III.2.3 – Variáveis categóricas

Considerando que as variáveis categóricas devem ser mutuamente exclusivas (PEDHAZUR, 1982), a variável categórica *Compatibilidade HLA* (E12) possui 34 receptores codificados com 3, 71 com código 2 e 4 receptores com código 1. A variável *Transfusões Sangüíneas* (E8) foi categorizada a partir dos quartis, sendo os maiores códigos aqueles que contribuem mais para o sucesso do transplante (SALOMÃO-FILHO *et al.*, 1986; GOLDFARB-RUMYANTZEV *et al.*, 2003). Assim, E8 foi codificada como:

- De 0 a 2 (inclusive) – código 4, 31 receptores;
- De 3 a 5 (inclusive) – código 3, 29 receptores;
- De 6 a 10 (inclusive) – código 2, 25 receptores;
- Acima de 11 – código 1, 24 receptores;

### III.3 – Padronização das variáveis

Todas as variáveis foram padronizadas para a faixa de variação 0,1-0,9, utilizando-se a Equação III.1 (ANDREA e KALAYEH, 1991):

$$Y_{esc} = 0,8 \cdot \frac{Y_i - Y_{min}}{Y_{max} - Y_{min}} + 0,1 \quad (III.1)$$

onde  $Y_{esc}$  é o valor padronizado,  $Y_i$  o valor da variável e  $[Y_{min}, Y_{max}]$  a faixa padronizada em que a variável de entrada se situa. O valor máximo das variáveis contínuas foi considerado como o de saturação (Tabela III.2).

A variável *Sobrevida do Enxerto* foi definida em cinco faixas, com resolução anual e tolerância de seis meses. Os resultados da conversão foram ajustados para o valor médio da faixa à qual pertencem (Ajuste de Faixa Escalada, em meses), padronizados segundo a Equação III.1.

**Tabela III.4** - Faixas (meses) de resolução anual da saída, com tolerância de 6 meses (0,08), para a Sobrevida do enxerto saturada em 60 meses. SE padronizado original é forçado para o centro da faixa. FE, AF e AFE significam Faixa da Escala, Ajuste da Faixa (valor médio da faixa, em meses) e Ajuste da Faixa na Escala (valor médio da faixa, na escala), respectivamente. Valores padronizados segundo a Equação III.1.

Faixa	FE	AF	AFE
[0-12]	0,10-0,26	6	0,18
]12-24]	0,26-0,42	18	0,34
]24-36]	0,42-0,58	30	0,50
]36-48]	0,58-0,74	42	0,66
]48-60]	0,74-0,90	54	0,82

### III.4 – Conjuntos de dados

Três conjuntos foram definidos, um para determinação dos parâmetros, outro para escolha do modelo e outro para teste do modelo (LJUNG, 1999; HAYKIN, 2002), aqui denominados de Conjunto de Estimação (CE), Conjunto de Validação (CV) e Conjunto de Teste (CT), respectivamente. No Apêndice B são mostrados os

diagramas de espalhamento das variáveis preditoras e da variável dependente, onde os testes de Kruskal-Wallis e  $\chi^2$  foram utilizados para teste de heterogeneidade entre os conjuntos.

Após o descarte dos casos com dados perdidos (60%, de 297 receptores), foi necessário fazer alocação aleatória estratificada, tentando evitar a introdução de heterogeneidades nos três conjuntos. A análise estratificada das variáveis foi então realizada, visando diminuir um possível viés de seleção. Assim, as variáveis *Rejeição* (E11) e *Transplantes Prévios* (E7) foram descartadas por não possuírem receptores nos três conjuntos. A quantidade final de variáveis preditoras foi 11. Após a identificação dos estratos, os conjuntos foram aleatoriamente divididos (Tabela III.5). As novas variáveis preditoras estão mostradas na Tabela III.6.

O Conjunto de Estimação teve dois receptores com *Compatibilidade HLA* de código 0 e o Conjunto de Validação e o Conjunto de Teste somente um receptor cada. Para a variável *Tipo de Diálise* cada um dos grupos foi contemplado com 1 paciente sob CAPD. Para a variável *Esquema Imunossupressor*, o valor 1 contribuiu com 41, 12 e 23 receptores para os Conjuntos de Estimação, Validação e Teste, respectivamente.

**Tabela III.5** - Número de receptores por variável estratificada, mantendo os três conjuntos com pelo menos um receptor por categoria nos três conjuntos.

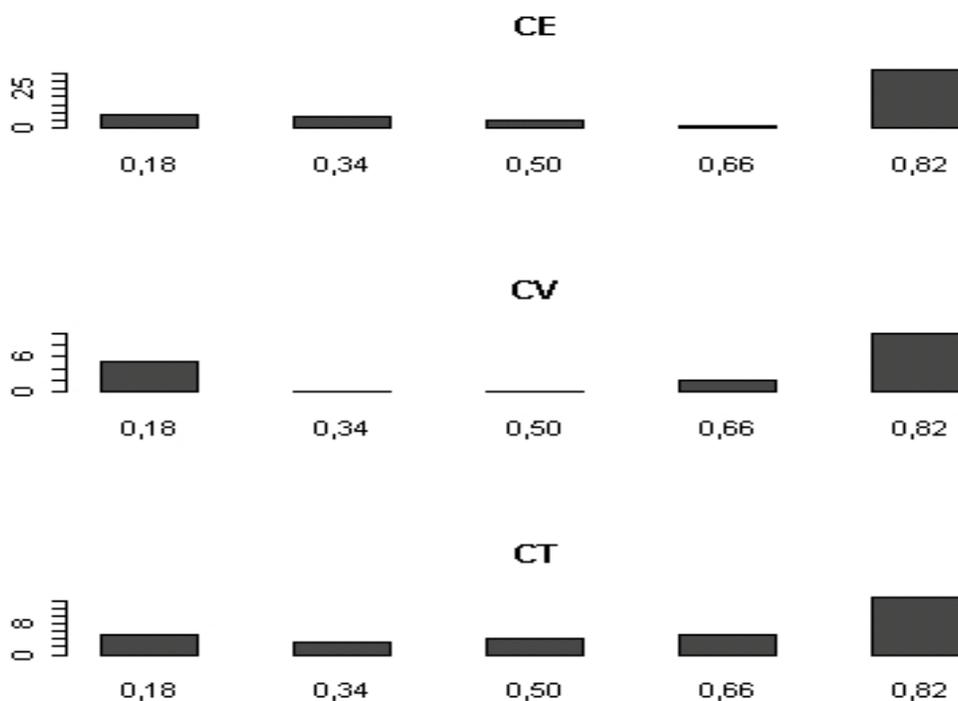
Conjuntos	E10 = 0	E10 = 1	E12 = 0	E12 = 0,5	E12 = 1
CE (60 receptores)	59	1	2	41	17
CV (17 receptores)	16	1	1	10	6
CT (32 receptores)	31	1	1	20	11
Total	106	3	4	71	34

**Tabela III.6**- Variáveis preditoras

Variáveis	Símbolo	Tipo da Variável
<i>Idade Receptor</i> (anos)	E1	Contínua
<i>Idade Doador</i> (anos)	E2	Contínua
<i>Sexo Receptor</i>	E3	Dicotômica, f = 1
<i>Sexo Doador</i>	E4	Dicotômica, f = 0
<i>Raça Receptor</i>	E5	Dicotômica, afro-descen = 0
<i>Raça Doador</i>	E6	Dicotômica, afro-descen = 0

<i>Transfusões Sangüíneas Receptor</i>	E8	Numérica
<i>Tempo de Diálise (meses)</i>	E9	Numérica
<i>Tipo da Diálise (HD, CAPD)</i>	E10	Dicotômica, HD = 0
<i>Compatibilidade HLA (Doador Vivo)</i>	E12	Categ, Id = 3, Haplo = 2, Não-Relacionado = 1
<i>Esquema Imunossupressor</i>	E13	Dicotômica, Uso de CSA = 1

Os Conjuntos de Estimação e Teste tiveram receptores em todas as faixas, porém o mesmo não aconteceu com o Conjunto de Validação. Somados os três conjuntos, as Faixas com valor médio 0,18 e 0,82 tiveram o maior número de casos, e estão presentes em todos os conjuntos. Devido ao menor número de receptores, o Conjunto de Validação não possui casos nas faixas 0,34 e 0,50. A distribuição dos valores de *Sobrevida do Enxerto* em cada conjunto está mostrada na Figura III.3.



**Figura III.3** - Distribuição das quantidades de receptores em cada faixa padronizada, por conjunto. As siglas CE, CV e CT significam Conjuntos de Estimação, Validação e Teste, respectivamente. O ponto central da faixa é mostrado na abscissa.

### III.5 – Discussão

A determinação de sobrevida baseada em dados estáticos pré-transplante é complexa (SHEPPARD *et al.*, 1999). Várias intercorrências podem advir logo após a cirurgia, e ainda, a constante introdução de novas estratégias de imunossupressão pode influenciar no comportamento do enxerto (SALOMÃO-FILHO *et al.*, 1996). Por exemplo, a utilização dos dados após a introdução da Ciclosporina A no protocolo imunossupressor, embora limitante do ponto de vista do número de casos completos, possibilitou, pelo próprio aumento da sobrevida proporcionada pela CsA, uma grande faixa de sobrevida (0-146 meses), o que poderia vir a favorecer a generalização pelo aumento do conteúdo de informação da base de dados, uma vez que se espera um aumento da sobrevida com a introdução de novos medicamentos imunossupressores.

Algumas variáveis têm sido consideradas importantes para a predição de sobrevida de enxerto renal, como a diferença entre creatinina a três e doze meses pós-TX (JOHNSON *et al.*, 2004), o estabelecimento de Função Renal Tardia (SHOSKES *et al.*, 1998), ou ainda o diagnóstico de Disfunção Crônica do Enxerto (MATSUI *et al.*, 2001) e a diferença da massa corporal entre doador e receptor (GOLDFARB-RUMYANTZEV *et al.*, 2003). RADEMAKER *et al.* (2003) sugerem a utilização do Índice de Resistência Arterial-Renal, obtido através ultrasonografia como preditor de perda do enxerto ou morte com enxerto funcionante. Outras variáveis preditoras são sugeridas em SIMIC-OGRIZOVIC *et al.* (1999).

No presente estudo, parte das variáveis pré-TX que poderiam ser incluídas foram descartadas devido à sua incompletude. Alguns métodos têm sido propostos para o preenchimento de variáveis com dados perdidos. Basicamente, existem três mecanismos de dados perdidos (*Missing Data Mechanism* – MDM): i) MCAR (*Missing Completely At Random*), no qual se assume que os dados perdidos são amostras aleatórias independentes do total de dados; ii) MAR (*Missing At Random*), no qual dados perdidos são independentes dos próprios dados perdidos, mas dependentes dos dados observados; iii) MNAR (*Missing Not At Random*), para o qual dados perdidos dependem dos próprios dados não observados (em oposição ao *Missing At Random*) (MENG, 2000; TWISK e de VENT, 2002; CLARK e ALTMAN, 2003).

Embora em grandes bases de dados seja comum a existência de dados perdidos, a abordagem de eliminá-los sistematicamente da análise pode levar a resultados tendenciosos, caso os dados excluídos não sejam *Missing Completely At Random*, que, entretanto, na prática não ocorre, pois usualmente o *Missing Data*

*Mechanism* é assumido como *Missing At Random* (MENG, 2000; CLARK e ALTMAN, 2003).

Segundo OOSTENBRINK *et al.* (2003), caso a prevalência seja superior a 50% de dados perdidos, a variável deve ser excluída da análise, pois poderia indicar dificuldade de obtenção do dado ou sua não importância à época do coleta. Por exemplo, a variável de sorologia positiva para o vírus da hepatite C só recentemente foi considerada, devido ao desenvolvimento de técnicas laboratoriais que possibilitaram sua determinação em larga escala.

Alguns dos métodos que têm sido considerados como válidos para o tratamento de dados perdidos, tais como o algoritmo EM (*Expectation-Maximization*, ver JOHNSON e WICHERN, 1988), assumem que os dados perdidos são *Missing At Random*, o que nem sempre ocorre (JOHNSON e WICHERN, 1988; MENG, 2000; TWISK e de VENT, 2002; CLARK e ALTMAN, 2003; OOSTENBRINK *et al.*, 2003). EFRON (1994) propõe três métodos utilizando *Bootstrap* na abordagem deste problema.

A abordagem de mineração dos dados e a extração de características da informação, inclusive quanto a distribuição estratificada, visou reduzir possível heterogeneidade entre os Conjuntos de Estimção, Validação e Teste, embora o pequeno número de casos completos ainda possa ter produzido resultados tendenciosos. A solução do problema de dados perdidos, discutido neste capítulo, ainda não se encontra resolvida na literatura. Além disso, deve-se reconhecer a complexidade do problema, embora a predição dicotômica, utilizada por diferentes autores, tais como DORSEY *et al.* (1997), SHEPPARD *et al.* (1999), sugira a viabilidade da aplicação de Redes Neurais Artificiais na predição de sobrevida de enxerto renal.

## CAPÍTULO IV

### Modelagem Linear Multivariável Aplicada a Transplante Renal

#### IV.1 – Introdução

Modelos matemáticos multivariáveis têm sido comumente utilizados na área biomédica, devido aos desfechos clínicos serem, freqüentemente, resultados de vários fatores, que podem, inclusive, estarem relacionados entre si. A modelagem tem sido empregada para determinação de causas, predição ou escolha de variáveis (HIRSH e RIEGELMAN, 1992; GISMONDI, 1999; GOLDFARB-RUMYANTZEV *et al.*, 2003; TERRIN *et al.*, 2003). Análise de regressão multivariável é utilizada em estatística para explicação de um fenômeno (resposta ou variável dependente) a partir de várias causas (variáveis preditoras ou independentes). O método clássico de análise de regressão linear é o Método de Mínimos Quadrados (MMQ – JOHNSON e WICHERN, 1992), utilizado na Regressão Linear Multivariável (RLM). Este é considerado um método robusto na predição, mesmo que alguns pressupostos não sejam satisfeitos para utilização em aplicações distintas (HAIR *et al.*, 1992 *apud* GISMONDI, 1999).

#### IV.2 –Regressão linear múltipla pelo método dos mínimos quadrados

As variáveis preditoras utilizadas, conforme Tabela IV.1, são contínuas (E1, E2, E8 e E9), dicotômicas (E3, E4, E5, E6, E10 e E13) e categórica (E12). Em modelos de regressão onde há variáveis categóricas entre as variáveis preditoras, pode-se adotar como procedimento (PEDHAZUR, 1982; HIRSH e RIEGELMAN, 1992):

- 1) Transformar a variável dependente em categórica;
- 2) Manter a variável dependente contínua e transformar a variável preditora categórica em dicotômicas mutuamente excludentes.

Visando a comparação com os resultados da rede neural, a opção escolhida foi a segunda. Assim, a variável categórica *Compatibilidade HLA* foi transformada em duas variáveis dicotômicas:

E14: *Compatibilidade HLA* Idêntico > sim (valor 0,9), não (valor 0,1);

E15: *Compatibilidade HLA* Haplo > sim (valor 0,9), não (valor 0,1);

Se E14 e E15 assumem valor 0,1, então o doador é não relacionado.

O modelo de regressão utilizado pode ser expresso por:

$$y_j = \hat{\beta}_0 \cdot x_{j0} + \hat{\beta}_1 \cdot x_{j1} + \hat{\beta}_{j2} \cdot x_{j2} + \dots + \hat{\beta}_{jr} \cdot x_{jr} + \hat{\epsilon}_j \quad (IV.1)$$

onde

$y_j$  é o valor verdadeiro de saída para o paciente j;

$\hat{\epsilon}_j$  é o resíduo entre o valor ajustado e o valor verdadeiro para o paciente j;

$\hat{\beta}_{jr}$  são os parâmetros a determinar, sendo r é o número de variável preditora, no caso  $r = 1, 2, \dots$ ;

$x_{jr}$  são as variáveis predictoras (regressoras) incluídas no modelo, sendo que usualmente  $x_{j0}$  assume o valor 1.

A expressão IV.1 pode ser re-escrita matricialmente como:

$$y = X \cdot \hat{b} + \hat{\epsilon} \quad (IV.2)$$

onde

$y$  é o vetor dos valores verdadeiros de saída;

$X$  é a matriz das variáveis regressoras;

$\hat{b}$  é o vetor dos parâmetros a determinar;

$\hat{\epsilon}$  é o vetor dos resíduos.

Para a determinação dos coeficientes das variáveis predictoras, utilizando-se o Método dos Mínimos Quadrados (MMQ), tem-se:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1r} \\ 1 & x_{21} & x_{22} & \dots & x_{2r} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nr} \end{bmatrix} \times \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_r \end{bmatrix} + \begin{bmatrix} \hat{\epsilon}_1 \\ \hat{\epsilon}_2 \\ \vdots \\ \hat{\epsilon}_n \end{bmatrix} \quad (IV.3)$$

ou então:

$$\hat{b} = (X^t \cdot X)^{-1} \cdot X^t \cdot y \quad (IV.4)$$

Logo, os valores ajustados são:

$$\hat{y} = \hat{H} \cdot y = X \cdot \hat{b} \quad (IV.5)$$

onde

$$\hat{H} = X \cdot (X^t \cdot X)^{-1} \cdot X^t \quad (IV.6)$$

e os resíduos em relação aos valores verdadeiros são:

$$E = (I - \hat{H}) \cdot y \quad (IV.7)$$

onde  $I$  é a matriz identidade e  $E$  é o vetor de resíduos (DAVINSON e HINKLEY, 1997).

Assumindo-se não haver dependência ou interação entre as variáveis, a equação IV.1 torna-se (PEDHAZUR, 1982):

$$y_j = \hat{\beta}_0 + \hat{\beta}_1.E1 + \hat{\beta}_2.E2 + \hat{\beta}_3.E3 + \hat{\beta}_4.E4 + \hat{\beta}_5.E5 + \hat{\beta}_6.E6 + \hat{\beta}_7.E8 + \hat{\beta}_8.E9 + \hat{\beta}_9.E10 + \hat{\beta}_{10}.E13 + \hat{\beta}_{11}.E14 + \hat{\beta}_{12}.E15 + \hat{\epsilon}_j \quad (IV.8)$$

pois a variável categórica E14 foi transformada em duas variáveis dicotômicas. A variável E8, categorizada, foi mantida como contínua para que o conteúdo da informação não se alterasse e, assim, melhorasse o desempenho do modelo.

### IV.3 – Índices de desempenho

O desempenho de modelos utilizando os Conjuntos de Estimação, Validação, e Teste pode ser avaliado segundo as seguintes expressões:

a) Proporção de Acertos nos Conjuntos:

$$AC = \frac{\text{Número de Acertos Dentro da Tolerância}}{\text{Número Total de Pacientes do Conjunto}} \quad (IV.9)$$

b) RMSE normalizado:

$$msen = \frac{1}{N} \times \sum_{i=1}^N \left( \frac{y_i - \hat{y}_i}{y_i} \right)^2 = \frac{1}{N} \times \sum_{i=1}^N \left( \frac{\hat{\epsilon}_i}{y_i} \right)^2 \quad (IV.10)$$

$$\therefore rmsen = \sqrt{msen}$$

c) Coeficiente de Determinação:

$$R^2 = \frac{\text{Variação Explicada}}{\text{Variação Total}} = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^N \hat{\epsilon}_i^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (IV.11)$$

onde:

$rmsen$  é a raiz quadrada do erro quadrático médio (do inglês *root mean square error*) normalizado. Sua unidade em relação à *Sobrevida do Enxerto* é percentual de meses padronizados;

$R^2$  é o Coeficiente de Determinação;

$N$  é o número de receptores do conjunto;

$y_i$  é o receptor de índice  $i$  do conjunto;

$\hat{y}_i$  é o valor ajustado para o receptor de índice  $i$ ;

$\bar{y}$  é a média dos valores de saída do conjunto;

$\hat{\epsilon}_i$  é o vetor de erros ajustados para o receptor  $i$ .

#### IV.4 – Aplicação da Regressão Linear Múltipla em transplante renal

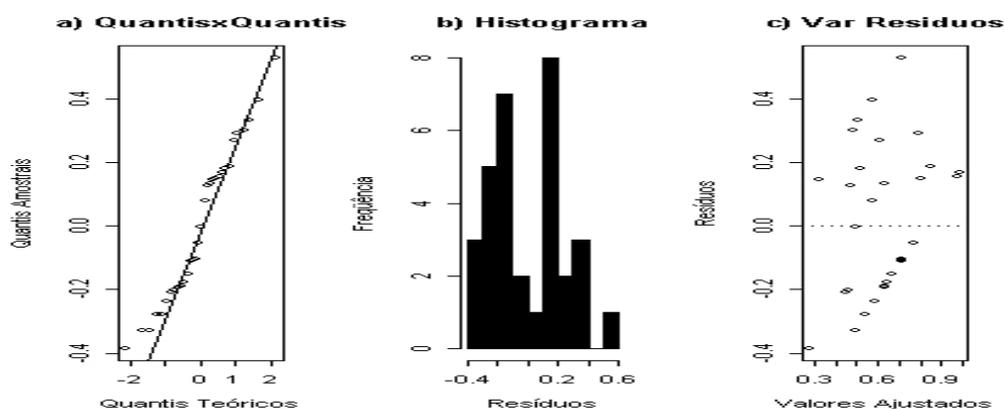
Os coeficientes de IV.8 foram estimados utilizando os pacientes dos Conjuntos de Estimação e de Validação (77 pacientes). A validação interna dos dados foi feita com o Conjunto de Teste (32 pacientes). O desempenho do modelo foi avaliado levando-se em consideração a proporção de acertos e o Coeficiente de Determinação Múltipla ( $R^2$ ) no Conjunto de Teste (JOHNSON e WICHERN, 1988; LJUNG, 1999).

O teste do modelo mostrado na equação IV.8 (utilizando o Conjunto de Teste) resultou em RMSE normalizado igual a 86%, proporção de acertos no Conjunto de Teste (ACT) de 2 em 32, Coeficiente de Determinação no Conjunto de Teste ( $R^2$ ) igual a 0,09. Os coeficientes da Equação IV.8 são mostrados na Tabela IV.1. Exceto pelos coeficientes relativos à Constante de Interseção e das variáveis *Idade do Receptor* e *Transfusões Sangüíneas*, os demais coeficientes resultaram em  $p$  superior ao nível de significância de 5%. Considerando somente os coeficientes para os quais se pode rejeitar a hipótese nula ( $H_0$ : coeficientes iguais a zero), as variáveis *Idade do Receptor* e *Transfusões Sangüíneas*, por terem coeficientes negativos, reduziriam a sobrevida do enxerto, ou seja, ambas estariam associadas negativamente com a sobrevida do enxerto.

**Tabela IV.1** - Coeficientes da reta de regressão dada pela equação V.3, identificados por MMQ, com a variável dependente *Sobrevida do Enxerto* saturada em 60 meses. A variável *Compatibilidade HLA* foi dicotomizada segundo item IV.2.

Variável Preditora	Sigla	Coef.	Valor - P
Constante de Interseção	-	1,1578	<0,001
<i>Idade Receptor</i>	<b>E1</b>	<b>-0,363</b>	<b>0,025</b>
<i>Idade Doador</i>	E2	0,090	0,497
<i>Sexo Receptor</i>	E3	-0,076	0,358
<i>Sexo Doador</i>	E4	-0,131	0,102

<i>Raça Receptor</i>	E5	-0,125	0,494
<i>Raça Doador</i>	E6	0,125	0,502
<b><i>Transf. Sangüíneas Receptor</i></b>	<b>E8</b>	<b>-0,287</b>	<b>0,018</b>
<i>Tempo de Diálise</i>	E9	-0,084	0,561
<i>Tipo da Diálise</i>	E10	-0,130	0,590
<i>Esquema Imunossupressor</i>	E13	0,162	0,165
<i>HLA Idêntico</i>	E14	-0,035	0,866
<i>HLA Haplo</i>	E15	-0,235	0,271

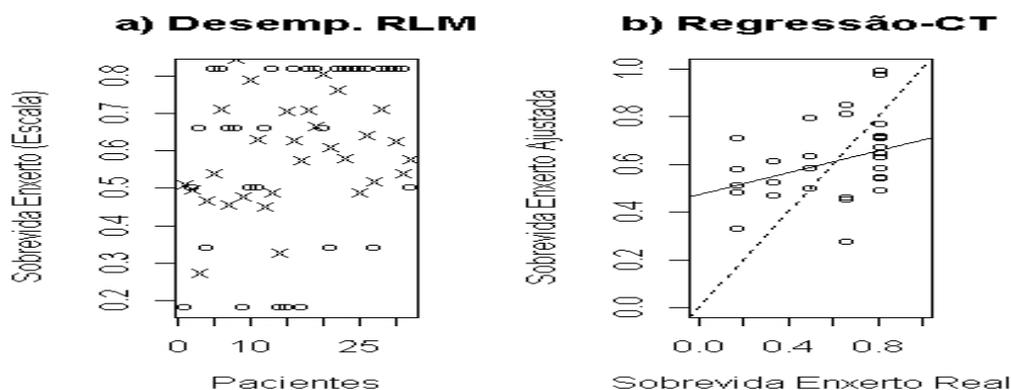


**Figura IV.1** - Gráficos de resíduos para RLM no CT. **a)** Diagrama *qq-plot* e **b)** Histograma **c)** Resíduos em função dos valores ajustados pela RLM

Análise por Shapiro-Wilk permitiu aceitar a hipótese nula de Gaussianidade para os resíduos no CT ( $W = 0,95$ ,  $p = 0,18$ ), sendo a média destes muito próxima de zero (0,004). No entanto, o modelo parece não preencher a suposição de homocedasticidade dos resíduos, pois estes não são constantes em relação aos seus valores ajustados (Figura IV.1-c) (JOHNSON e WICHERN, 1988; GISMONDI, 1999; LJUNG, 1999; AGUIRRE, 2002).

Os gráficos de desempenho do modelo de Regressão Linear Múltipla estão sumarizados na Figura IV.2. A reta de regressão ajustada do modelo RLM aplicada ao CT ( $y = 0,23x + 0,47$ ) é mostrada na Figura IV.2-b, bem como a reta de regressão ideal. A reta de regressão ideal tem erro de predição nulo para todos os pontos da resposta da rede, e, portanto, tem coeficiente angular unitário. Na Figura IV.2-a podem ser vistos os pontos de *Sobrevida do Enxerto* reais e seus valores ajustados pela

Regressão Linear Múltipla. Estão dentro da tolerância estipulada os receptores 2 e 22 (Tabela IV.2). Os pontos mostrados sobre a reta ideal na Figura IV.2-b são os receptores cuja predição foi obtida dentro da tolerância. Os demais representam os receptores dentro das faixas correspondentes.



**Figura IV.2** - Desempenho da RLM no CT, mostrando **a)** comparação entre os pontos de sobrevida real (o) e ajustada pela RLM (x), **b)** a reta de regressão ideal (linha tracejada) onde os pontos se encontrariam se a correlação fosse perfeita, e a calculada (linha contínua) entre a sobrevida real e a ajustada.

**Tabela IV.2** – Comparação entre os valores encontrados para o modelo RLM com saturação da sobrevida em 60 meses (estimação dos parâmetros utilizando CE + CV) e os valores de sobrevida real para os receptores do Conjunto de Teste. O receptor de numero 32 praticamente ficou dentro da tolerância (0,08, equivalente a seis meses).

Rec.	Real	RLM	Rec	Real	RLM60	Rec	Real	RLM60
#1	0,18	0,511	#12	0,66	0,453	#23	0,82	0,583
<b>#2</b>	<b>0,50</b>	<b>0,499</b>	#13	0,82	0,492	#24	0,82	0,989
#3	0,66	0,283	#14	0,82	0,328	#25	0,82	0,491
#4	0,34	0,468	#15	0,18	0,710	#26	0,82	0,645
#5	0,82	0,541	#16	0,82	0,630	#27	0,34	0,521
#6	0,82	0,715	#17	0,18	0,577	#28	0,82	0,715
#7	0,66	0,459	#18	0,82	0,710	#29	0,82	0,975
#8	0,66	0,848	#19	0,82	0,669	#30	0,82	0,628
#9	0,18	0,481	#20	0,66	0,810	#31	0,82	0,543
#10	0,50	0,793	#21	0,34	0,611	#32	0,50	0,581
#11	0,50	0,634	<b>#22</b>	<b>0,82</b>	<b>0,766</b>			

## IV.5 – Considerações sobre a utilização de RLM em transplantes

Modelos de predição multivariável lineares são os mais utilizados na literatura. Neste trabalho utilizou-se a Regressão Linear Múltipla pelo Método dos Mínimos Quadrados por se tratar do método mais simples e possuir robustez para predição, mesmo quando não estão satisfeitos os pressupostos de sua utilização, como pdf Gaussiana (Normal) dos resíduos, com variância constante e média zero. No presente trabalho, os resultados utilizando este tipo de modelagem apresentaram Coeficiente de Determinação de 9%, indicando fraca associação linear entre as variáveis preditoras e a dependente, sugerindo a utilização de modelos não lineares para este problema. Outro aspecto a ser considerado é que a equação de regressão foi ajustada para ter erro quadrático médio dos dados apresentados para a estimação do modelo (com o Conjunto de Estimação mais o Conjunto de Validação), e que, portanto, sua capacidade preditiva teve que ser avaliada utilizando dados distintos (Conjunto de Teste).

Para o modelo de regressão linear determinado com o conjunto de 109 pacientes com transplante de doadores vivos, somente as variáveis *Idade do Receptor*, *Transfusões Sanguíneas* e a constante de Interseção resultaram em coeficientes com nível de significância inferior a 5%, estando tais variáveis associadas negativamente com a *Sobrevida do Enxerto*. GOLDFARB-RUMYANTZEV *et al.* (2003) também relatam este tipo de associação entre a sobrevivida do enxerto e a idade dos receptores (superior a 45 anos), porém para doadores cadavéricos. Constatação similar é feita por SALOMÃO-FILHO *et al.* (1986), quanto ao numero de transfusões.

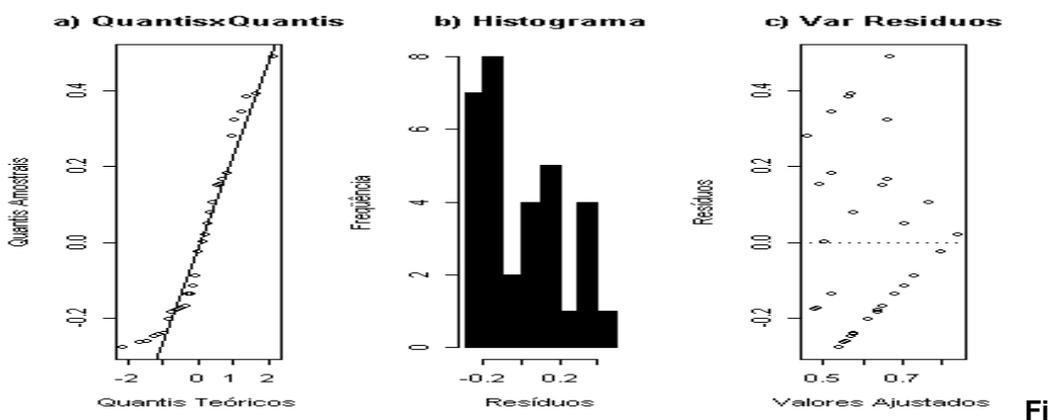
Caso se considerasse somente a *Idade do Receptor* e as *Transfusões Sanguíneas*, o modelo RLM resultante seria:

$$SE = 0,890 - 0,279 \times E1 - 0,216 \times E8 \quad (IV.12)$$

onde *SE* é a *Sobrevida do Enxerto*; *E1* é a variável *Idade do Receptor* e *E8* as *Transfusões Sanguíneas*. Utilizando-se neste modelo o Conjunto de Teste, a proporção de acertos seria de 5 em 32, com RMSE normalizado de 89,1% e Coeficiente de Determinação de 0,11. Logo, estes resultados diferem em muito daquele do modelo RLM completo no que concerne a proporção de acertos (2 em 32), tendo, porém, valores de RMSE (86%) e do Coeficiente de Determinação (0,09) bastante próximos. As características dos resíduos para o Conjunto de Teste, conforme mostrado na Figura IV.3, não atendem aos pressupostos da utilização do

Método de Mínimo Quadrados, como Gaussianidade (Shapiro-Wilk,  $p = 0,01$ ) e homocedasticidade. Além disso, as retas de regressão ajustada para o Conjunto de Teste ( $y = 0,14x + 0,52$ ) e de regressão ideal são mostradas na Figura IV.4-b. Os cinco receptores para os quais a tolerância de seis meses pode ser observada são mostrados em negrito na Tabela IV.3.

Em ambos os modelos RLM, o completo e o com somente as variáveis *Idade do Receptor* e *Transfusões Sangüíneas*, fica evidenciada a tendenciosidade, pois, os pressupostos para a utilização do método não são satisfeitos. Assim, os resultados sugerem a necessidade de se investigar modelagem não-linear, com vistas a melhoria do desempenho na predição da *Sobrevida do Enxerto*.

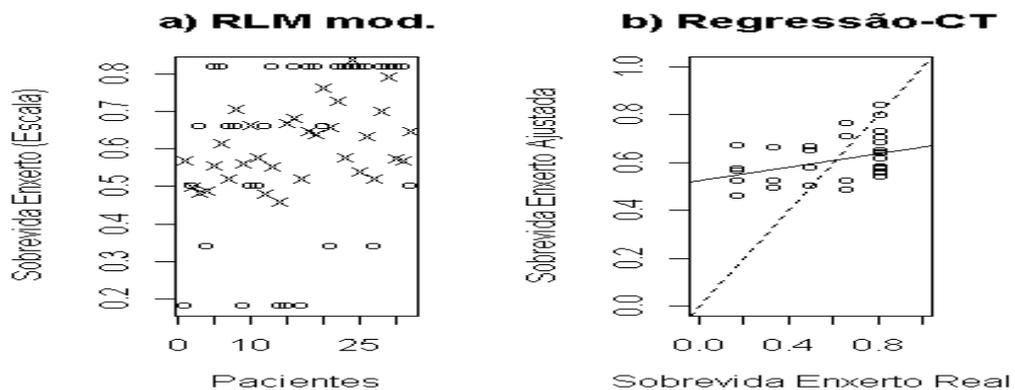


**Figura IV.3** - Gráficos de resíduos para RLM modificado no CT. **a)** Diagrama *qq-plot* **b)** Histograma **c)** Resíduos em função dos valores ajustados pelo RLM modificado.

**Tabela IV.3** – Comparação entre os valores encontrados para o modelo RLM modificado com saturação da sobrevida em 60 meses (estimação dos parâmetros utilizando CE + CV) e os valores de sobrevida real para os receptores do Conjunto de Teste.

Rec.	Real	RLM	Rec	Real	RLM	Rec	Real	RLM
#1	0,180	0,573	#12	0,660	0,483	#23	0,820	0,581
#2	<b>0,500</b>	<b>0,503</b>	#13	0,820	0,556	<b>#24</b>	<b>0,820</b>	<b>0,840</b>
#3	0,660	0,486	#14	0,820	0,461	#25	0,820	0,542
#4	0,340	0,492	#15	0,180	0,670	#26	0,820	0,636
#5	0,820	0,559	#16	0,820	0,684	#27	0,340	0,522
#6	0,820	0,617	#17	0,180	0,522	#28	0,820	0,703
#7	0,660	0,522	#18	0,820	0,650	<b>#29</b>	<b>0,820</b>	<b>0,795</b>

#8	0,660	0,709	#19	0,820	0,642	#30	0,820	0,578
#9	0,180	0,564	#20	0,660	0,765	#31	0,820	0,573
#10	0,500	0,665	#21	0,340	0,662	#32	0,500	0,650
#11	0,500	0,578	#22	0,820	0,731			



**Figura IV.4** - Desempenho da RLM modificada no CT, mostrando a) comparação entre os pontos de sobrevida real (o) e ajustada pela RLM (x), b) a reta de regressão ideal (linha tracejada) onde os pontos se encontrariam se a correlação fosse perfeita, e a calculada (linha contínua) entre a sobrevida real e a ajustada.

# CAPÍTULO V

## Modelagem por RNA em Transplante Renal

### V.1 – Introdução

As Redes Neurais Artificiais (RNA) têm tido crescente aceitação em problemas que envolvem predição na área biomédica (DVORCHIK *et al.*, 1996; DORSEY *et al.*, 1997; SHEPPARD *et al.*, 1999; HUGHES *et al.*, 2001; GISMONDI *et al.*, 2002; DAS *et al.*, 2003), constituindo uma alternativa não-linear aos métodos lineares clássicos, por não necessitarem dos pressupostos de Gaussianidade ou linearidade, necessários à modelagem por Regressão Linear Múltipla, por exemplo (ENNET *et al.*, 2001, HAYKIN, 2002; DAS *et al.*, 2003). O tipo de RNA utilizado neste trabalho (*Perceptron* Multi-Camada - PMC) foi escolhido por ser esta a RNA mais utilizada nesta área (vide Capítulo III).

Na área de transplantes, não existem muitos artigos publicados utilizando modelagem por Redes Neurais Artificiais em predição, principalmente quando a variável dependente (saída) não é dicotômica, conforme visto no Capítulo III. Assim, este trabalho pretendeu desenvolver metodologias para ajudar no dimensionamento de uma rede neural utilizada para predição cuja saída seja contínua e com poucos dados para treinamento.

### V.2 - Definição do número de replicações Bootstrap (BST)

Para a determinação do número mínimo de amostras *Bootstrap* necessárias para replicação do Conjunto de Estimção, a Estatística de Interesse (EI) escolhida foi o maior autovalor da matriz Normal ( $X^t.X$ ) (AGUIRRE, 2000), onde X é a matriz composta de 60 pacientes do Conjunto de Estimção (linhas) e as 11 variáveis de entrada (colunas). O programa LabVIEW foi utilizado para este processamento.

Embora o problema de predição em sobrevida em Transplantes possa ser classificado como um problema de “Aproximação de Funções”, no qual não há pressupostos de linearidade, a escolha recaiu nesta estatística linear para uma fácil caracterização da estrutura da base de dados original. A EI encontrada para a amostra original foi 178,32. Os passos do procedimento foram:

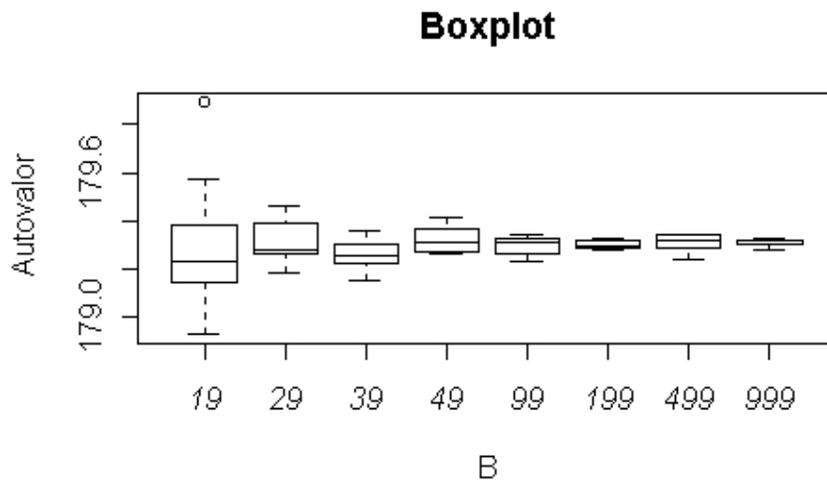
- a) Primeiramente, definiu-se valores B de replicações BST para a matriz Normal (19, 29, 39, 49, 99, 199, 499 e 999), os quais foram comparados para a identificação do número mínimo de replicações que manteria a mesma média e variância de B = 999. Para isto, utilizou-se a reamostragem *Bootstrap* balanceada dos pacientes (BSTbalp);
- b) Cada paciente foi replicado B vezes, formando uma nova matriz de B x 60 linhas e 11 colunas;
- c) Com a ajuda de um gerador de números pseudo-aleatórios, cada linha foi escolhida uma única vez para a formação de B matrizes, cada qual com 60 linhas e 11 colunas;
- d) Cada uma destas matrizes foi multiplicada pela sua transposta e determinado o maior autovalor (EI) da matriz resultante;
- e) Calculou-se o valor médio da EI dentre as B amostras;
- f) Este procedimento foi repetido 10 vezes, e então, foram determinadas a média e a variância desses dez valores, para cada B especificado.

O gerador de números pseudo-aleatórios utilizado possui função densidade de probabilidade uniforme. O programa pode ter como entrada qualquer um dos conjuntos de receptores (Conjunto de Estimação, Conjunto de Validação ou Conjunto de Teste). Assim, os conjuntos são matrizes de 60, 17 ou 32 linhas (receptores) e 11 colunas (as onze variáveis preditoras E1, E2, E3, E4, E5, E6, E8, E9, E10, E12 e E13 encontram-se descritas no Capítulo III). Além disso, o programa tem a flexibilidade do número de replicações *Bootstrap* ser estabelecido pelo usuário. Os novos pares de Entrada-Saída assim gerados são então armazenados em formato que facilita a utilização por programas desenvolvidos em MatLab e R.

Os maiores autovalores obtidos estão mostrados na Tabela V.1, sendo os *boxplots* relativos a cada grupo mostrados na Figura V.1. As análises por Kruskal-Wallis (tendência central) e Fligner-Killeen (variabilidade) dos oito grupos mostraram resultados distintos. Enquanto o primeiro não indicou diferenças significativas (KW = 8,42, p = 0,30), o segundo mostrou resultado contrário (FK = 28,2, p = 0,0002). A aplicação do teste de Fligner-Killeen para B variando de 99 a 999 não indicou haver diferença entre as variâncias (FK = 2,70, p = 0,44). Logo, o valor escolhido de B para BSTbalp foi 99, sendo o mesmo utilizado para o treinamento da RNA com o Conjunto de Estimação.

**Tabela V.1** - Maiores autovalores encontrados para cada replicação, utilizando BST balanceado dos receptores. Em cada uma das dez simulações foi calculada a média entre as B amostras BST.

B	19	29	39	49	99	199	499	999
1	179,9	179,3	179,2	179,4	179,3	179,3	179,3	179,3
2	178,9	179,4	179,2	179,3	179,3	179,3	179,3	179,3
3	179,2	179,4	179,3	179,3	179,3	179,3	179,3	179,3
4	179,3	179,3	179,3	179,3	179,2	179,3	179,2	179,3
5	179,1	179,3	179,2	179,3	179,3	179,3	179,3	179,3
6	179,2	179,3	179,4	179,3	179,3	179,3	179,3	179,3
7	179,4	179,2	179,4	179,3	179,3	179,3	179,3	179,3
8	179,1	179,2	179,2	179,4	179,2	179,3	179,3	179,3
9	179,3	179,3	179,2	179,4	179,3	179,3	179,3	179,3
10	179,6	179,5	179,3	179,3	179,3	179,3	179,3	179,3
média	179,3	179,3	179,3	179,3	179,3	179,3	179,3	179,3
variância	0,0725	0,0083	0,0041	0,0030	0,0016	0,0003	0,0013	0,0002



**Figura V.1** - *Boxplots* dos autovalores para B = 19 até B = 999.

### **V.3 – Aplicação de RNA em transplante renal**

O desenvolvimento da Rede Neural Artificial foi baseado em DEMUTH e BEALE (1998), SHEPPARD *et al.* (1999), HAYKIN (2002) e TERRIN *et al.* (2003). Os passos a seguir esclarecem a metodologia utilizada.

#### **V.3.1 – Inicialização de pesos**

O método Nguyen-Widrow (DEMUTH e BEALE, 1998; YAM e CHOW, 2000) acelera o treinamento ajustando os pesos iniciais espaçados sobre a faixa de operação dos neurônios e foi o escolhido para o treinamento das redes. Como os pesos podem assumir valores negativos (YAM e CHOW, 2000), as funções de ativação são do tipo tangente hiperbólica, com faixa de variação entre  $-1$  e  $+1$  para neurônios da camada oculta. A função de ativação do neurônio de saída é log-sigmoidal e possui faixa de operação  $0 - 1$ , uma vez que a variável preditora somente assume valores positivos.

#### **V.3.2 – Seleção da arquitetura da RNA**

A escolha da arquitetura da Rede Neural Artificial foi feita utilizando-se *Bootstrap* no Conjunto de Estimção, o método *Split Sample* (LJUNG, 1999; HAYKIN, 2002; TERRIN *et al.*, 2003) e Parada Antecipada (DEMUTH e BEALE, 1998; HAYKIN, 2002). Arquiteturas com maior número de neurônios ocultos foram comparadas com as de menor, até que os índices de desempenho fossem estatisticamente iguais, excluindo-se as de maior complexidade (maior quantidade de neurônios). Os testes não-paramétricos de Kruskal-Wallis e Wilcoxon-Mann-Whitney (HIRSH e RIEGELMAN, 1992) para amostras não-pareadas foram usadas para a comparação estatística (estatísticas KW e U,  $H_0$ : igualdade de tendências centrais). Os testes de Fligner-Killeen (PARK e LINDSAY, 1999) e Ansari-Bradley foram usados para comparação de variabilidade (estatísticas FK e AB,  $H_0$ : igualdade de variâncias) e o de Shapiro-Wilk (LOONEY, 1995) para verificação da hipótese de Gaussianidade. A comparação estatística para mesma arquitetura foi feita pelo teste de Wilcoxon pareado (estatística V,  $H_0$ : igualdade de tendências centrais). Caso o valor  $P$  correspondente à estatística calculada fosse maior do que o nível de significância, a hipótese nula era aceita. Em todos os testes o nível de significância  $\alpha$  foi de 0,05.

As Arquiteturas testadas possuíam somente uma camada oculta e foram denotadas por *ExOxS*. Neste trabalho, utilizou-se número de entradas (variáveis preditoras)  $E = 11$ , número de neurônios de saída  $S = 1$  e  $O$  número de neurônios da camada oculta. A Rede Neural Artificial era do tipo *Feedforward*, totalmente conectada, treinada com algoritmo de retropropagação de erro. O algoritmo de treinamento foi o gradiente descendente com taxa de aprendizagem adaptativa e constante de momento *default* do MatLab, versão 5.3. A apresentação dos pares de Entrada-Saída (ES) foi por lote, portanto, cada época (atualização dos pesos) foi computada após a apresentação de todos os receptores do conjunto. Cada sessão de treinamento utilizou pares ES em diferentes seqüências.

O índice de desempenho para definição das arquiteturas foi estabelecido como a proporção de acertos no conjunto de dados, e definido como a razão entre a soma dos acertos da rede sobre o total de pacientes do conjunto. Cada acerto foi definido como *saída da RNA dentro da tolerância de seis meses em relação à saída original*, para cada paciente. O melhor modelo foi selecionado pela proporção de acertos no Conjunto de Validação (ACV), pois este possui dados distintos do Conjunto de Estimação, utilizado para a estimação dos parâmetros livres da rede (LJUNG, 1999; HAYKIN, 2002). A validação interna dos dados (teste do modelo) foi feita pela verificação da proporção de acertos no Conjunto de Teste (ACT).

### **V.3.2.1 - Determinação do número de neurônios ocultos**

Como mencionado, considerando não haver regra geral para a seleção de arquiteturas, inicialmente definiu-se uma faixa de valores para o número de neurônios ocultos mais abrangente, e, progressivamente, diminuiu-se a complexidade da rede. Outra heurística utilizada para diminuir o excesso de ajuste foi a Parada Antecipada (DEMUTH e BEALE, 1998; HAYKIN, 2002). A Parada Antecipada foi realizada utilizando o Conjunto de Validação e as etapas para a determinação da faixa inicial do número de neurônios ocultos foram:

- i) Foram realizadas dez sessões de treinamento com neurônios ocultos variando de 50 a 200% do número de entradas da rede. Para  $E = 11$ , testou-se  $O = 5, 10, 15$  e  $20$ ;
- ii) A arquitetura de melhor desempenho entre quatro utilizadas (considerada como a Arquitetura-Base, com número de neurônios ocultos  $O_{base}$ ) foi a escolhida para o “refinamento” da definição da arquitetura final;

iii) Uma vez estabelecida a Arquitetura-Base em ii), a faixa final de neurônios ocultos escolhida para seleção da melhor arquitetura foi [Obase-2, Obase-1, Obase, Obase+1, Obase+2, Obase+3].

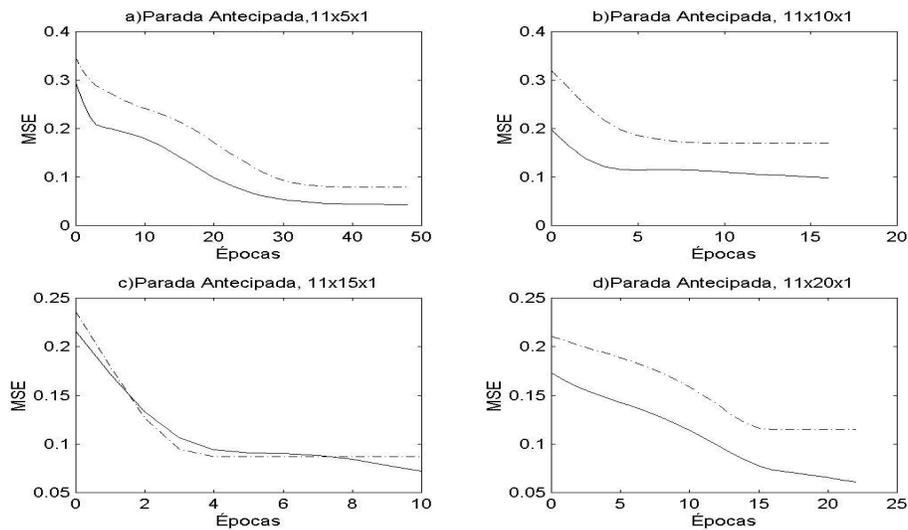
Os resultados referem-se a uma faixa de variação das entradas [0,1-0,9], variáveis contínuas com saturação no 95º percentil. A saída tinha escala com resolução de 12 meses, e “tolerância” de seis meses (ou seja, a saída era considerada “certa” caso se localizasse até seis meses acima ou abaixo do valor real – valor padronizado igual a 0,080 - ver Tabela III.4).

Os resultados para a proporção de acertos no Conjunto de Estimação (ACE) e no Conjunto de Teste (ACT) após dez sessões de treinamento são mostrados na Tabela V.2. A análise por Kruskal-Wallis e Fligner-Killeen não mostrou diferença significativa entre as arquiteturas com O = 5, 10, 15 e 20 para ACT (KW = 1,70, p = 0,64 e FK = 3,98, p = 0,26). Os desempenhos das RNA com Parada Antecipada são mostrados na Figura V.2. Como a hipótese de haver diferença não pode ser aceita, a Arquitetura Base escolhida foi a de cinco neurônios na camada oculta. Para O = 5, foram escolhidas as arquiteturas 11x3x1, 11x4x1, 11x5x1, 11x6x1, 11x7x1 e 11x8x1 para “refinamento” da arquitetura mais adequada. Toda a modelagem foi realizada pelo programa MatLab 5.3.

**Tabela V.2** – Proporção de acertos no CE (em 60 receptores) e no CT (em 32 receptores), para treinamento e validação utilizando Parada Antecipada.

Sessão	ACE5	ACE10	ACE15	ACE20	ACT5	ACT10	ACT15	ACT20
1	8/60	16/60	16/60	8/60	4/32	6/32	8/32	6/32
2	15/60	17/60	23/60	21/60	9/32	10/32	6/32	6/32
3	17/60	13/60	17/60	12/60	7/32	7/32	7/32	5/32
4	12/60	12/60	10/60	13/60	6/32	7/32	10/32	9/32
5	12/60	21/60	16/60	19/60	7/32	6/32	13/32	10/32
6	9/60	5/60	15/60	20/60	6/32	5/32	7/32	7/32
7	17/60	19/60	16/60	21/60	7/32	4/32	1/32	13/32
8	12/60	20/60	20/60	20/60	9/32	8/32	11/32	8/32
9	14/60	25/60	11/60	14/60	6/32	6/32	7/32	11/32
10	13/60	21/60	14/60	13/60	4/32	8/32	5/32	1/32
Média	12,9/60	16,9/60	15,8/60	16,1/60	6,5/32	6,7/32	7,5/32	7,6/32

ACE e ACT significam proporção de acertos nos Conjuntos de Estimação e Teste. Os algarismos posteriores às siglas são o número de neurônios na camada oculta de cada RNA



**Figura V.2-** Gráficos mostrando a evolução do treinamento, com parada antecipada utilizando o CV. 3ª sessão. MSE: *Mean Square Error*. Linha tracejada: decaimento do MSE para o CV. Linha contínua: decaimento do MSE para o CE.

### V.3.2.2 - Eliminação das arquiteturas com menor proporção de acertos para o Conjunto de Estimação

Para as arquiteturas 11x3x1, 11x4x1, 11x5x1, 11x6x1, 11x7x1 e 11x8x1, estabeleceu-se que aquelas que não obtivessem desempenho satisfatório no treinamento (resultados estatisticamente inferiores dos encontrados para ACE da arquitetura 11x8x1), seriam descartadas. Dez Sessões de Treinamento *Bootstrap* (STB) foram realizadas, cada uma com 99 amostras *Bootstrap* do Conjunto de Estimação (CEBST) mais o Conjunto de Estimação original. Após o treinamento, as variáveis de entrada dos Conjuntos de Estimação e Validação foram aplicadas à rede, e seus resultados comparados com as sobrevidas dos enxertos reais (originais) de cada um dos conjuntos. Com este procedimento, esperou-se eliminar as arquiteturas com menor ACE, com (ACEBST) e sem *Bootstrap*. Como índices de desempenho foram empregados o ACE e o ACEBST, sendo as etapas de avaliação as seguintes:

- i) Terminada a sessão de treinamento com o Conjunto de Estimação original, a mesma rede foi utilizada para o treinamento das 99 amostras

CEBST. Tal procedimento visa manter todas as amostras com as mesmas condições iniciais. Obtém-se ACE e ACEBST;

- ii) Repetiu-se i) 10 vezes;
- iii) Foram determinadas as arquiteturas semelhantes sem *Bootstrap* pelo ACE;
- iv) Idem com BST pelo ACEBST;
- v) Foram comparados os grupos com e sem *Bootstrap* separadamente.

Com vistas a evitar valores de ACE muito baixos, optou-se por aumentar o tempo de treinamento para 500 épocas ou MSE = 0,01 (o que ocorrer primeiro), de modo a se ter um mínimo local “menos raso”, pois, no procedimento de parada antecipada, observou-se valores de ACE inferiores 10/60 (máximo das médias de 11,3/60).

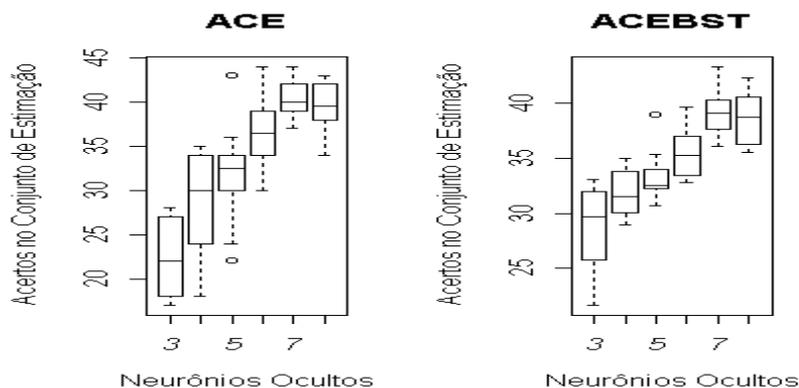
As sessões de treinamento *Bootstrap* para O = 3, 4, 5, 6, 7 e 8 apresentaram os resultados da Tabela V.3 para o Conjunto de Estimação (ACE e ACEBST), e da Tabela V.4 (ACV e ACVBST) para o Conjunto de Validação. Pelo teste de Kruskal-Wallis verifica-se haver diferenças significativas entre os seis grupos para ACE (KW = 0,47,  $p < 0,0001$ ). Entretanto, retirando-se as arquiteturas 11x3x1, 11x4x1, 11x5x1 e 11x6x1 não foram encontradas diferenças significativas (KW = 0,58,  $p = 0,44$ ), sendo estas arquiteturas eliminadas para ACE. Resultados similares foram encontrados para ACEBST (KW = 43,62,  $p < 0001$ ) para as seis arquiteturas, mas não para as arquiteturas com 7 e 8 neurônios na camada oculta (KW = 0,14,  $p = 0,71$ ). O *boxplot* para cada arquitetura é mostrado na Figura V.3. A avaliação das redes foi também realizada com O = 7 e 8 (com e sem BST).

**Tabela V.3** – Média *Bootstrap* para a proporção de acertos no CE (ACEBST, em 60 receptores) para O = 3, 4, 5, 6, 7 e 8.

,Sessão	ACE3	ACE4	ACE5	ACE6	ACE7	ACE8
1	30,8/60 (26/60)	33,8/60 (34/60)	32,5/60 (33/60)	32,8/60 (30/60)	37,5/60 (38/60)	39,9/60 (41/60)
2	30,6/60 (18/60)	33,5/60 (35/60)	39,0/60 (43/60)	32,8/60 (30/60)	40,3/60 (44/60)	36,3/60 (36/60)
3	32,8/60 (27/60)	28,9/60 (25/60)	32,6/60 (34/60)	35,8/60 (39/60)	43,4/60 (44/60)	36,8/60 (39/60)
4	33,0/60 (22/60)	30,0/60 (23/60)	35,4/60 (34/60)	36,0/60 (37/60)	38,5/60 (39/60)	40,6/60 (42/60)
5	21,6/60 (18/60)	29,3/60 (31/60)	31,6/60 (30/60)	34,0/60 (34/60)	37,7/60 (39/60)	42,3/60 (43/60)
6	23,6/60 (21/60)	31,4/60 (29/60)	30,8/60 (24/60)	37,0/60 (38/60)	39,3/60 (40/60)	36,3/60 (38/60)

7	28,9/60 (27/60)	35,0/60 (34/60)	32,4/60 (31/60)	34,7/60 (36/60)	39,0/60 (40/60)	41,4/60 (43/60)
8	32,0/60 (28/60)	31,7/60 (18/60)	33,3/60 (32/60)	39,7/60 (44/60)	40,7/60 (42/60)	37,8/60 (38/60)
9	25,8/60 (17/60)	31,0/60 (24/60)	32,2/60 (22/60)	38,4/60 (39/60)	36,2/60 (37/60)	39,7/60 (40/60)
10	27,0/60 (22/60)	33,9/60 (35/60)	34,0/60 (36/60)	33,5/60 (34/60)	39,8/60 (42/60)	35,5/60 (34/60)
Média	28,6/60 (22,6/60)	31,9/60 (28,8/60)	33,4/60 (31,9/60)	35,5/60 (36,1/60)	39,2/60 (40,5/60)	38,7/60 (39,4/60)

O numeral posterior às siglas: quantidade de neurônios na camada oculta. Abaixo e entre parênteses está ACE, referentes ao treinamento com o CE original.



**Figura V.3** - *Boxplots* mostrando a distribuição da proporção de acertos do CE original (ACE) e com BST (ACEBST) para treinamento das redes com  $O = 3, 4, 5, 6, 7$  e  $8$  neurônios na camada oculta.

**Tabela V.4** - Média *Bootstrap* da proporção de acertos para o Conjunto de Validação (ACVBST, 17 receptores).

Sessão	ACV3	ACV4	ACV5	ACV6	ACV7	ACV8
1	6,1/17 (5/17)	8,2/17 (3/17)	5,7/17 (2/17)	6,8/17 (5/17)	4,4/17 (2/17)	6,0/17 (4/17)
2	4,9/17 (2/17)	5,6/17 (1/17)	7,1/17 (4/17)	5,5/17 (3/17)	5,5/17 (3/17)	5,6/17 (3/17)
3	7,1/17 (4/17)	6,1/17 (5/17)	5,6/17 (2/17)	5,5/17 (3/17)	5,8/17 (3/17)	8,2/17 (6/17)
4	7,7/17 (3/17)	5,8/17 (3/17)	7,1/17 (4/17)	4,9/17 (3/17)	6,7/17 (4/17)	5,7/17 (3/17)
5	4,7/17 (4/17)	4,8/17 (1/17)	6,4/17 (2/17)	5,4/17 (2/17)	9,4/17 (8/17)	6,8/17 (4/17)

6	4,2/17 (4/17)	6,8/17 (5/17)	6,6/17 (4/17)	10,0/17 (8/17)	5,0/17 (2/17)	6,4/17 (4/17)
7	6,4/17 (5/17)	5,6/17 (1/17)	6,0/17 (5/17)	6,8/17 (6/17)	6,6/17 (5/17)	6,7/17 (4/17)
8	7,3/17 (5/17)	7,0/17 (3/17)	5,7/17 (3/17)	6,4/17 (4/17)	6,3/17 (3/17)	6,0/17 (3/17)
9	5,9/17 (3/17)	6,2/17 (2/17)	5,6/17 (2/17)	5,7/17 (3/17)	4,7/17 (3/17)	7,4/17 (5/17)
10	4,3/17 (2/17)	6,1/17 (2/17)	6,0/17 (4/17)	6,0/17 (5/17)	5,9/17 (2/17)	7,8/17 (5/17)
média	5,9/17 (3,7/17)	6,2/17 (2,6/17)	6,2/17 (3,2/17)	6,3/17 (4,2/17)	6,0/17 (3,5/17)	6,7/17 (4,1/17)

Numeral posterior às siglas: quantidade de neurônios na camada oculta. Abaixo e entre parênteses estão as quantidades dos acertos referentes ao treinamento com o CE original (ACV).

### V.3.2.3 - Verificação da proporção de acertos com e sem BST

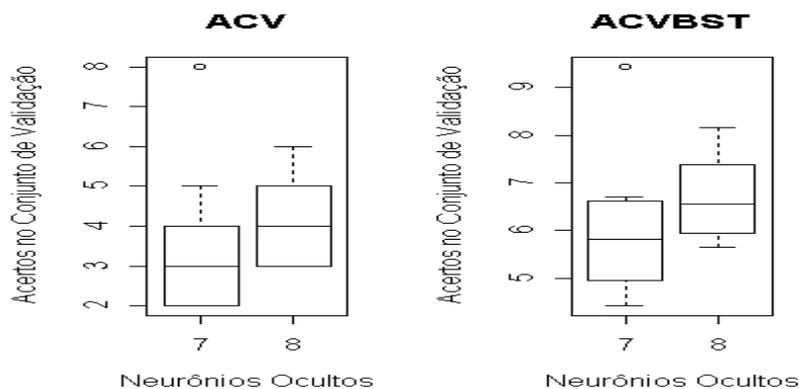
A determinação da arquitetura de menor número de parâmetros, utilizando *Bootstrap* ou não, foi realizada comparando-se o desempenho das arquiteturas 11x7x1 e 11x8x1, para ACV e ACVBST. Os testes de Wilcoxon-Mann-Whitney (para tendência central) e de Ansari-Bradley (para variabilidade) foram utilizados na comparação. Não foram encontradas diferenças significativas entre as arquiteturas 11x7x1 e 11x8x1 para ACVBST ( $U = 29$ ,  $p = 0,11$ ) e tampouco para ACV ( $U = 29$ ,  $p = 0,12$ ). Os resultados para homogeneidade de variância também não apresentaram diferenças estatisticamente significativas (Tabela V.5 e Figura V.4). Para ACV obteve-se  $AB = 47$ ,  $p = 0,16$ , e para ACVBST os resultados foram  $AB = 50$  e  $p = 0,50$ .

Uma vez que não foram encontradas diferenças entre os grupos, a arquitetura escolhida com e sem *Bootstrap* foi 11x7x1. A análise por Wilcoxon pareado mostrou diferenças estatisticamente significativas para a utilização de *Bootstrap* do Conjunto de Estimação em relação ao treinamento somente com o Conjunto de Estimação original ( $V = 0$ ,  $p = 0,002$ ). A utilização de 99 replicações BST apresentou maiores valores para ACV (Tabela V.5 e Figura V.4).

**Tabela V.5** - Média *Bootstrap* da proporção de acertos para o Conjunto de Validação (ACVBST, 17 receptores) das arquiteturas selecionadas com maior proporção de acertos no CE.

Sessão	ACV7	ACV8
1	4,4/17 (2/17)	6,0/17 (4/17)
2	5,5/17 (3/17)	5,6/17 (3/17)
3	5,8/17 (3/17)	8,2/17 (6/17)
4	6,7/17 (4/17)	5,7/17 (3/17)
5	9,4/17 (8/17)	6,8/17 (4/17)
6	5,0/17 (2/17)	6,4/17 (4/17)
7	6,6/17 (5/17)	6,7/17 (4/17)
8	6,3/17 (3/17)	6,0/17 (3/17)
9	4,7/17 (3/17)	7,4/17 (5/17)
10	5,9/17 (2/17)	7,8/17 (5/17)
Média	6,0/17 (3,5/17)	6,7/17 (4,1/17)

Numeral posterior às siglas: quantidade de neurônios na camada oculta. Abaixo e entre parênteses estão as proporções dos acertos referentes ao treinamento sem BST



**Figura V.4** -- *Boxplots* mostrando a distribuição da proporção de acertos do Conjunto de Validação sem BST (ACV) e com BST (ACVBST).

#### V.4 – Verificação da generalização

A arquitetura 11x7x1 foi testada quanto a generalização (validação interna) utilizando *Bootstrap* e comparada com o treinamento utilizando o Conjunto de Estimção original. O critério de parada das Sessões de Treinamento *Bootstrap* (STrBST) para a determinação da melhor configuração de pesos iniciais foi ACV igual ao número inteiro imediatamente superior à média de ACV (igual a 4) e ACVBST (igual a 6), ou mil sessões de treinamento (o que ocorresse primeiro). Ao fim destas, foram registrados o ACV com o ACT correspondente (Tabela V.6), o RMSE normalizado em relação ao Conjunto de Teste (RMSEnt) e os Coeficientes Angulares e de Interseção da reta de regressão. Este procedimento foi repetido dez vezes. Para teste do modelo, o Coeficiente de Determinação no Conjunto de Teste também foi registrado. A melhor configuração de pesos foi considerada a que proporcionou o maior ACT dentre os dez ensaios. O critério escolhido visou evitar a superespecialização da rede para o Conjunto de Validação. O mesmo procedimento foi conduzido para a média encontrada para ACVBST (Tabela V.7). Os resultados para a rede 11x7x1 (ACT) comparados pelo teste de Wilcoxon pareado não mostraram diferenças estatisticamente significativas ( $V = 24,5$ ,  $p = 0,86$ ).

**Tabela V.6** – Aplicação da rede 11x7x1 no Conjunto de Treinamento (32 receptores), para ACV mínimo igual a 4. RMSEN, M, e B são o RMSE normalizado em CT, o Coeficiente Angular da reta de regressão, Intercepto da reta de regressão, respectivamente. Rede treinada com 60 receptores.

STrBST	ACV	ACT	RMSEN	M	B	R
1	4/17	9/32	105	0,26	0,50	0,27
2	5/17	9/32	90,2	0,31	0,48	0,31
3	6/17	7/32	115	0,14	0,59	0,13
4	4/17	7/32	72,1	0,36	0,39	0,41
5	5/17	6/32	76,0	0,26	0,43	0,25
6	4/17	3/32	98,5	0,26	0,50	0,28
7	6/17	7/32	110	0,19	0,58	0,21
8	4/17	7/32	117	0,02	0,64	0,03
<b>9</b>	<b>4/17</b>	<b>11/32</b>	<b>84,0</b>	<b>0,49</b>	<b>0,33</b>	<b>0,47</b>
10	5/17	7/32	121	0,02	0,61	0,02

Em negrito encontram-se os resultados proporcionados pelo maior ACT

**Tabela V.7** – Aplicação da rede 11x7x1 no Conjunto de Treinamento (32 receptores), para ACVBST mínimo igual a 6. RMSEN, M, e B são o RMSE normalizado no CT, o Coeficiente Angular da reta de regressão, o Intercepto da reta de regressão e o Coeficiente de Determinação, respectivamente. Rede treinada com 60 receptores.

STrBST	ACVBST	ACTBST	RMSEN	M	B	R
1	6/17	7/32	88,2	0,27	0,49	0,28
2	7/17	6/32	72,5	0,43	0,33	0,42
3	6/17	3/32	113	-0,01	0,68	-0,01
4	6/17	7/32	115	0,10	0,57	0,09
5	6/17	7/32	65,1	0,45	0,38	0,44
6	7/17	9/32	121	0,01	0,63	0,01
7	6/17	8/32	58,9	0,42	0,34	0,38
8	6/17	8/32	93,3	0,39	0,45	0,38
9	6/17	7/32	101	0,18	0,54	0,21
10	8/17	10/32	100	0,25	0,50	0,25

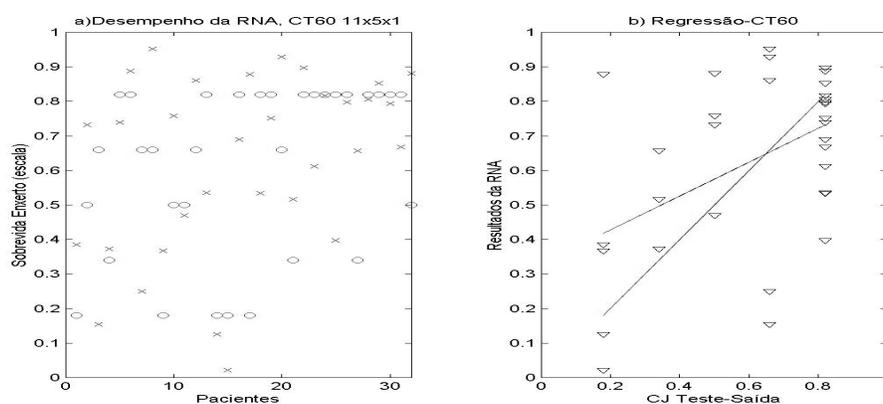
## V.5 – Desempenho da RNA selecionada

Os resultados da arquitetura 11x7x1 (que proporcionou ACT7 = 11, Tabela V.6) são mostrados na Tabela V.8. Os índices de desempenho ACT, RMSE normalizado do Conjunto de Teste e o coeficiente de determinação  $R^2$  (equação IV.11) para a rede foi comparado com o modelo de Regressão Linear Múltipla.

As retas de regressão, ideal e ajustada, entre a resposta da RNA para cada receptor e os valores de *Sobrevida do Enxerto* reais, para o Conjunto de Teste, são mostrados na Fig. V.5-b. A reta de regressão ideal tem erro de predição nulo para todos os pontos da resposta da rede, e, portanto, tem coeficiente angular unitário. A reta de regressão ajustada da RNA ( $y = 0,49x + 0,33$ ) indica uma pequena correlação linear entre as variáveis independentes e dependente. A comparação entre os valores reais de *Sobrevida do Enxerto* com os valores ajustados pela RNA, conforme mostra a Figura V.5-a, indica que, para os receptores 4, 6, 11, 14, 19, 22, 24, 26, 28, 29 e 30, os valores da RNA estão muito próximos dos valores reais (Tabelas III.4 e V.9). Os pontos mostrados sobre a reta ideal são os receptores cuja predição situou-se dentro da tolerância de seis meses (0,08 padronizados). Os demais pontos representam os receptores dentro das faixas correspondentes. Os valores encontrados estão mostrados na Tabela V.9.

**Tabela V.8** - Comparação dos resultados entre RLM e a RNA. RLM: transformação da variável categórica *Compatibilidade HLA* e 77 receptores.

Índices de Desempenho	RLM	RNA
Parâmetros	13	92
ACT (Nº acertos/Pacientes do CT)	2/32	11/32
RMSEnt (%)	86,0	84,0
R <sup>2</sup> t	0,09	0,47



**Figura V.5** - Desempenho da RNA (com CE no treinamento) em CT, mostrando **a)** comparação entre os pontos de sobrevida real (o) e ajustada pela RNA (x), **b)** a reta de regressão ideal (coeficiente angular de 45°, i.e, correlação perfeita), e a regressão linear determinada com os valores de sobrevida reais e os resultados da rede.

**Tabela V.9** – Comparação entre os valores encontrados para a RNA e os valores de sobrevida reais para os receptores do Conjunto de Teste .

Rec.	meses	AFE	RNA	Rec	meses	AFE	RNA	Rec	meses	AFE	RNA
#1	6	0,180	0,385	#12	42	0,660	0,860	#23	54	0,820	0,612
#2	30	0,500	0,732	#13	54	0,820	0,536	#24	54	<b>0,820</b>	<b>0,817</b>
#3	42	0,660	0,154	#14	6	<b>0,180</b>	<b>0,126</b>	#25	54	0,820	0,398
#4	18	<b>0,340</b>	<b>0,373</b>	#15	6	0,180	0,022	#26	54	<b>0,820</b>	<b>0,798</b>
#5	54	0,820	0,373	#16	54	0,820	0,690	#27	18	0,340	0,658
#6	54	<b>0,820</b>	<b>0,888</b>	#17	6	0,180	0,878	#28	54	<b>0,820</b>	<b>0,807</b>
#7	42	0,660	0,250	#18	54	0,820	0,534	#29	54	<b>0,820</b>	<b>0,853</b>
#8	42	0,660	0,953	#19	54	<b>0,820</b>	<b>0,751</b>	#30	54	<b>0,820</b>	<b>0,794</b>

#9	6	0,180	0,368	#20	42	0,660	0,929	#31	54	0,820	0,668
#10	30	0,500	0,758	#21	18	0,340	0,516	#32	30	0,500	0,881
#11	30	<b>0,500</b>	<b>0,470</b>	#22	54	<b>0,820</b>	<b>0,898</b>				

Em negrito, receptores cujos resultados ajustados da rede se encontram dentro da tolerância. AFE é o Ajuste de Faixa na Escala (padronizada).

## V.6 – Discussão

Um dos problemas de modelagem estatística na área de saúde é a dificuldade em se obter um tamanho amostral adequado, pois, muitas vezes, a prevalência de uma doença é pequena ou a disponibilidade dos dados insuficiente, dificultando a generalização dos resultados (ALMEIDA *et al.*, 2001; CHANG *et al.*, 2002; GISMONDI *et al.*, 2002, INFANTOSI *et al.*, 2004). Se o número de elementos de uma amostra é pequeno e a distribuição estatística é desconhecida, métodos computacionalmente intensivos tais como o *Bootstrap* balanceado podem ser utilizados para a estimação e inferência de parâmetros da população, permitindo aumentar o poder do teste estatístico e a acurácia dos estimadores (para maiores detalhes, ver Apêndice C).

Neste trabalho, utilizou-se um banco de dados contendo 297 pacientes que receberam transplante de doadores vivos a partir de 1986, quando da introdução da Ciclosporina A como droga imunossupressora. Com vistas ao uso de dados pré-TX na predição de sobrevida do enxerto renal por Redes Neurais Artificiais, a “mineração” dos dados foi previamente realizada (Capítulo III) e, então, a arquitetura da RNA foi investigada usando-se *Bootstrap* e Validação Cruzada. As técnicas aplicadas também objetivaram determinar a Rede Neural Artificial de menor dimensão e, assim, aumentar a aceitação de modelos de predição do tipo “Aproximador de Funções”, ou do tipo “Caixa-Preta” (LJUNG, 1999). A metodologia descrita neste trabalho tem a potencialidade de se constituir em alternativa de treinamento/estimação de Redes Neurais Artificiais para aplicação em bases de dados com características distintas da de transplantados renais, aqui investigada. Caso isto seja confirmado, esta metodologia poderá ser utilizada como ferramenta de auxílio à decisão médica.

Modelos de predição similares ao investigado neste trabalho, porém para aplicação distinta (Tabela V.8), também utilizaram saídas dicotômicas, o que evidencia a dificuldade da predição na forma de categorias ou dados contínuos. SHEPPARD *et al.* (1999) relatam que preferiram estudar a ocorrência de citomegalovirose após Transplante Renal do que a sobrevida de enxerto renal, face à complexidade da

predição de sobrevida do enxerto. Na Tabela V.8 nota-se a grande variação no número de variáveis de entrada e de neurônios ocultos utilizados, o que poderia ser atribuído tanto à complexidade do problema em estudo, quanto à possível inexistência de metodologia para o estabelecimento da arquitetura da rede. Deve-se ainda enfatizar que o grande número de variáveis preditoras, com provável interação entre as mesmas, e as observações perdidas (*missing data*) quando da coleta de dados também implicariam no aumento da complexidade do problema.

**Tabela V.8** – Trabalhos selecionados na área de transplantes (vide Capítulo II), utilizando RNA em predição. Número de receptores divididos em dados de treinamento/dados de teste. As siglas VCDUF, F.A., Gauss, Log, Heav, Dic e ND significam Validação Cruzada Deixe-Um-de-Fora, Função de Ativação dos neurônios, Gaussiana, sigmoidal, *Heaviside*, dicotômica e não disponível, respectivamente.

AUTOR	ÓRGÃO	Nº Rec	Alg. Trein.	Arquitetura	F.A.
DORSEY <i>et al.</i>	Pâncreas	132/28	BP	12x20x1	Gauss
SHEPPARD <i>et al.</i>	CMV-Rim	VCDUF	BP	17x10x2 42x20x2	Log
FURNESS <i>et al.</i>	RejAg-Rim	125/21	ND	12x1	Hard
MELVIN <i>et al.</i>	TXH	64/31	BP	4x5x1 8x10x1 9x11x1	Log
SHOSKES <i>et al.</i>	TXR	100/20	ND	12x30x10x2	ND

A utilização de técnicas de simulação computacionalmente intensiva (*Bootstrap*) foi parte de uma estratégia de seleção do modelo de melhor generalização. A hipótese é que, ao se introduzir amostras *Bootstrap* no treinamento, seriam criadas outras amostras com as mesmas características estatísticas da amostra original e, como as Redes Neurais Artificiais utilizadas (*Perceptron* Multi-Camada com funções de ativação sigmoidais) se comportam como “máquinas aproximadoras de funções”, isto possibilitaria a seleção do modelo com melhor predição dos dados de teste. Entretanto, a hipótese não se verificou, mas indicou um caminho para o desenvolvimento de metodologias de seleção de arquiteturas mais explícitas. O *Perceptron* Multi-Camada foi o tipo de Rede Neural Artificial escolhido para este trabalho devido ao objetivo de desenvolver metodologia para predição de variáveis

contínuas, além de ser o tipo de RNA mais comumente empregado em aplicações similares.

Duas opções para desenvolvimento futuro emergem do trabalho em relação ao uso de *Bootstrap*. Primeiro, a definição do número de replicações escolhida (99) pode ser re-estudada, com a escolha de outra Estatística de Interesse e, assim, procurar verificar se, com outro valor para o número de replicações, haveria alteração na escolha de melhor arquitetura. Segundo, diferentemente do usado no presente trabalho, no qual usou-se BST balanceado, a reamostragem de algumas variáveis pode ser refeita (porém, ainda de modo, a não se “gerar pacientes com características impossíveis”) com vistas a aumentar o conteúdo do conjunto de treinamento.

A definição de um número mínimo de replicações de *Bootstrap* balanceadas é importante, para evitar os longos tempos de simulação computacional requeridos para as replicações usualmente requeridas quando é utilizado o *Bootstrap* tradicional (vide Apêndice C). Em trabalho recentemente publicado, HAHN *et al.* (2005) também utilizaram o *Bootstrap* no treinamento de uma Rede Neural Artificial (simulação de um conjunto de dados de treinamento com  $B = 20, 40$  pacientes), porém sem discutir os resultados com *Bootstrap* em comparação a situação de não aplicação desta técnica.

A determinação da quantidade de neurônios na camada oculta não é trivial. Se por um lado o modelo com maior número de parâmetros pode ser capaz de solucionar um determinado problema, por outro, não se pode garantir que tal solução proporcione uma generalização adequada, uma vez que um elevado número de neurônios na camada oculta aumenta a possibilidade de super-especialização da rede (REED, 1993; DEMUTH e BEALE, 1998; HAYKIN, 2002). Alternativamente, como empregado no presente trabalho, a arquitetura pode ser determinada testando-se várias configurações de camada oculta e, então, verificando-se qual a de melhor desempenho (REED, 1993).

A metodologia utilizada para a definição da arquitetura da Rede Neural Artificial partiu de uma faixa abrangente de número de neurônios ocultos, definindo uma Arquitetura Base, e, a partir desta, uma faixa mais restrita. Isto possibilitou a definição da arquitetura 11x7x1 para dados originais e para os dados utilizando *Bootstrap* no treinamento. Na literatura analisada (Capítulo II), não foi possível identificar a utilização de testes estatísticos para a escolha da arquitetura, mas, somente, a utilização de índices de desempenho como número de épocas, área sob a curva ROC ou MSE. Neste trabalho, a metodologia aplicada à escolha das arquiteturas (pelo menos 10 sessões de treinamento com pesos iniciais aleatórios, para cada arquitetura,

testando-se  $H_0$ : igualdade entre grupos em relação à tendência central e variabilidade) foi menos “Caixa-Preta” do que algoritmos automáticos de “poda” de rede ou Validação Cruzada. Assim, a escolha da melhor arquitetura levou em consideração as condições de menor número de parâmetros, com a média da proporção de acertos no Conjunto de Validação (ACV). Essa é uma heurística de “poda” de redes e visa principalmente à escolha da arquitetura de menor complexidade, evitando-se o excesso de ajuste (REED, 1993; HAYKIN, 2002).

A estratégia de generalização usada, com dados do Conjunto de Teste oriundos da mesma população que os de treinamento, segue HAYKIN (2002), mas difere do conceito de generalização utilizado em epidemiologia (validação externa), no qual a mesma é comprovada utilizando-se dados de populações distintas (TERRIN *et al.*, 2003; CROSS *et al.*, 2003; DAS *et al.*, 2003; STEYERBERGER *et al.*, 2003). Assim, o conceito de generalização aqui utilizado deve ser entendido como validação interna dos dados (utilizando conjunto de dados distintos do conjunto usado na estimação dos parâmetros, porém ambos oriundos da mesma população), e não como validação externa.

Tendo em vista a utilização de neurônios sigmoidais, há a possibilidade de saturação de alguns nós da rede se a escolha dos pesos iniciais for puramente aleatória, aumentando o tempo de treinamento. Assim, a escolha do vetor de pesos iniciais por técnicas estatísticas visa à diminuição do tempo de treinamento (DEMUTH e BEALE, 1998; YAM e CHOW, 2000) e vários algoritmos têm sido desenvolvidos para este fim. O utilizado neste trabalho foi o Nguyen-Widrow, e DEMUTH e BEALE (1998) sugerem ser este algoritmo adequado a neurônios sigmoidais. A utilização de Taxa de Aprendizagem adaptativa foi utilizada por aumentar a velocidade de convergência no treinamento, assim como também a utilização de constante de momento (JACOBS, 1988; DEMUTH e BEALE, 1998).

Durante o treinamento, verificou-se que menores valores da proporção de acertos no Conjunto de Validação associavam-se com as maiores proporções de acertos no Conjunto de Teste (dados não mostrados). Tal constatação poderia ser interpretada, talvez, como se a busca pela melhor configuração de pesos resultasse em uma superespecialização com Conjunto de Validação, dificultando, assim, a generalização da RNA. Embora estudos adicionais devam ser conduzidos para investigar esta hipótese, é de ser esperar que, devido à homogeneidade entre os Conjuntos de Validação e Teste, os percentuais de acertos sejam semelhante.

O modelo de Rede Neural desenvolvido neste trabalho resultou em maior Coeficiente de Determinação (0,47) em comparação com o da modelagem por Regressão Linear Múltipla (0,09). Além disso, para a RNA a proporção de acertos (34,4%) é muito superior ao obtido pela RLM (6,2%), mesmo utilizando um menor número de exemplos para estimação dos parâmetros. Dos 11 receptores do Conjunto de Teste, cuja predição encontrava-se dentro da tolerância padronizada de 0,08 (6 meses), 8 situavam-se na faixa padronizada de valor médio igual a 0,82 (de 48 a 60 meses), talvez pelo fato de que grande parte dos receptores terem tido *Sobrevida do Enxerto* superior a 60 meses e, como a Função de Ativação do neurônio de saída é log-sigmoidal, pode ter havido saturação. Conforme exposto, os resultados obtidos com a RNA sugerem a potencialidade do emprego da metodologia desenvolvida.

## CAPÍTULO VI

### Conclusão

Dos pacientes com doadores cadavéricos, metade perderá os enxertos em até 10 anos (MARSDEN, 2003). Tal afirmativa evidencia que a predição de sobrevida de enxertos renais é um problema relevante, tanto do ponto de vista clínico quanto do aspecto ético, e indica a complexidade do processo de tomada de decisão quanto à realização de transplante renal.

De uma base de dados original contendo 297 receptores com doadores vivos, devido à sua incompletude, e, após aplicação de técnica de mineração de dados, somente 109 pacientes foram selecionados para este estudo de predição de sobrevida de enxerto renal (neste caso, devido a disponibilidade dos dados, a predição foi limitada a 60 meses). O conjunto em estudo, mesmo de dimensão reduzida, teve que ser subdividido (conjuntos de Estimação, Validação e Teste), com vistas ao desenvolvimento e teste dos modelos. Além disso, sempre que possível, procurou-se manter a homogeneidade dos conjuntos de modo a se obter uma melhor generalização dos modelos.

Onze variáveis preditoras (*Idade do Doador e Receptor, Sexo do Doador e Receptor, Raça do Doador e Receptor, Duração Prévia da Diálise, Tipo da Diálise, Transfusões Sangüíneas, Compatibilidade HLA e Esquema Imunossupressor*) foram selecionadas, com base na opinião de nefrologista especialista em transplante (responsável na instituição em que os receptores tem sido, ou foram, acompanhados). Como tais variáveis têm sido freqüentemente citadas na literatura especializada, considerou-se desnecessária a investigação da sensibilidade das mesmas. Indiscutivelmente, a utilização de variáveis pós-Transplante imediato poderia melhorar a predição da *Sobrevida do Enxerto*, bem como outras variáveis pré-Transplante poderiam ter sido incluídas, não o tendo sido devido a ausência de dados. Além disso, o aumento do número de preditores, embora possa resultar em melhora no conteúdo de informação, implica em aumento de dimensionalidade e, conseqüentemente, da complexidade do problema, podendo dificultar a generalização do modelo resultante.

A modelagem clássica (Regressão Linear Múltipla, RLM) resultou em um percentual de acertos no Conjunto de Teste de 6,2%, raiz do Erro Médio Quadrático normalizado de 86% (equivalente a 10 meses) e Coeficiente de Determinação de 9%. As variáveis preditoras com coeficientes significativos (valor  $p < 0,05$ ) foram *Idade do Receptor* e *Transfusões Sangüíneas*. Tais resultados indicam que a RLM parece não

ter sido capaz de detectar a complexidade do processo, o que talvez pudesse ocorrer usando-se modelagem não-linear, baseado em Rede Neural Artificial (RNA).

Assim, uma RNA 11x7x1, totalmente conectada, treinada com algoritmo de retropropagação do erro (baseado no método de Gradiente Descendente, com taxa de aprendizagem adaptativa e constante de momento), com neurônios utilizando funções de ativação sigmoidais foi adotada visando prever a sobrevida do enxerto em escala contínua. Para a determinação do número de neurônios na camada oculta desenvolveu-se metodologia que utiliza Validação Cruzada, *Bootstrap* (BST) e testes estatísticos não-paramétricos, inicialmente, para arquiteturas com maior número de parâmetros até alcançar um modelo de menor complexidade com desempenho estatisticamente similar. A Rede Neural assim determinada resultou em 34,4% de acerto para o Conjunto de Teste, além de o Coeficiente de Determinação de 47% e de raiz do erro médio quadrático de 84%. Este percentual de acerto para a RNA é 5,5 vezes maior do que o obtido com o modelo de Regressão Linear Múltipla pelo Método dos Mínimos Quadrados (6,2%).

Cabe ainda salientar que a metodologia adotada, além de ser distinta das heurísticas tradicionais, é mais compreensível do que, por exemplo, algoritmos genéticos ou técnicas de “poda” baseadas na matriz Hessiana. As hipóteses nulas para a determinação do número de neurônios ocultos eram do tipo “*Os grupos analisados são iguais, dado o nível de significância de 5%?*”, o que pode amenizar a “Síndrome da Caixa-Preta”, citada por CROSS *et al.* (2003) e diminuir a resistência à utilização de sistemas com Redes Neurais Artificiais na prática clínica. O *Bootstrap* balanceado, empregado para aumentar a informação contida no treinamento da Rede Neural Artificial, é um método já estabelecido para inferência estatística de pequenas amostras e, portanto, de direto entendimento. Entretanto, atestando a complexidade na predição com variável dependente contínua, a aplicação da RNA selecionada, embora tenha mostrado desempenho superior ao modelo de regressão linear, mostra a necessidade de um aprofundamento das técnicas empregadas neste trabalho.

A fim de melhorar a predição em escala contínua ou categorizada, sugere-se:

- a) Definição do número de replicações *Bootstrap* necessário para a seleção da arquitetura que forneça uma maior generalização. Pode ser estudado o desempenho de várias replicações a partir de uma única Rede Neural Artificial, por exemplo, a 11x7x1, e, então, a partir da comparação com os resultados obtidos, definir-se um outro B aplicável ao caso;

- b) A utilização de *Bootstrap* em variáveis escolhidas, de forma a não haver incompatibilidade entre o doador e receptor (por exemplo, Tempo de diálise igual a zero e Hemodiálise para tipo de diálise).
- c) Implementação de uma metodologia para preenchimento dos dados perdidos, por exemplo, utilizando métodos tradicionais (algoritmo *Expectation-Maximization* ou Imputação Múltipla) ou *Bootstrap*;
- d) A definição de cinco faixas de *Sobrevida do Enxerto* implicou, na prática, em categorizar-se a variável dependente. Embora o *Perceptron* Multi-Camada com algoritmo de retropropagação seja considerado um aproximador universal de funções (desde que as funções de ativação sejam diferenciáveis e contínuas), não foi idealizada uma rede classificadora de padrões neste trabalho. A utilização de um sistema híbrido de várias redes ou um *Perceptron* Multi-Camada com cinco neurônios de saída pode ser estudada;
- e) A utilização de Funções de Ativação no neurônio de saída que não saturem, por exemplo, a função exponencial ou linear. O treinamento com Regularização Bayesiana parece ser aplicável a este caso, onde há poucos exemplos de treinamento;
- f) A utilização de outras técnicas tais como análise de Curva de Sobrevida ou modelo de Azares Proporcionais de Cox, além de Regressão Não-Linear, para comparação do desempenho podem ser implementadas, dependendo da característica da variável dependente;
- g) Aumento no tempo de seguimento, para verificar a sobrevida dos receptores transplantados mais recentemente. Também a inclusão da data do transplante pode ser incluída, tendo em vista que o aumento da experiência da equipe transplantadora e a adoção de novos medicamentos imunossupressores influencia positivamente na sobrevida do enxerto (Ávila *et al.*, 2004)

Espera-se, a partir dos resultados obtidos, aplicar as metodologias desenvolvidas em uma base de dados de transplantados renais de outro centro transplantador, para validação externa do modelo de Rede Neural Artificial selecionado, na predição de sobrevida em enxertos renais com variável dependente contínua.

## CAPÍTULO VII

### Referências

AGUIRRE, L. A., 2000, *Identificação de Sistemas – Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais*, Belo Horizonte, Editora UFMG.

ALMEIDA, R. M. V. R., INFANTOSI, A. F. C., GISMONDI, R., 2001, “Replicação Bootstrap e Análise de Sensibilidade em Redes Neurais Artificiais” In: *Proceedings of the V Brazilian Conference on Neural Networks*, Rio de Janeiro, RJ, Brazil, pp. 295-300, April 2-5.

ANDREA, T. A., KALAYEH, H., 1991, “Applications of Neural Networks in Quantitative Structure-Activity Relationships of Dihydrofolate Reductase Inhibitors”, *Journal of Medical Chemistry.*, v.34, pp. 2824-2836.

ÁVILA, C. A., L., RUZANY, F., SOUZA, E. R. M., SAMPAIO, J. C., DORIGO, D., SUASSUNA, J. H. R., 2004, “Long-Term Outcome of Renal Transplantation at a University Center in Rio de Janeiro Between 1975 and 2000”, *Transplantation Proceedings*, v. 35, n. 4, pp. 865-867.

BABA, N., MOGAMI, Y., KOHZAKI, M., SHIRAISHI, Y., YOSHIDA, Y., 1994, “A Hybrid Algorithm for Finding the Global Minimum of Error Function of Neural Networks and Its Applications”, *Neural Networks*, v. 7, n. 8, pp. 1253-1265.

BILLINGS, S. A., JAMALUDDIN, H. B., CHEN, S., 1992, “Properties of Neural Networks With Applications to Modeling Non-Linear Dynamical Systems”, *International Journal of Control*, v. 55, n. 1, pp. 193-224.

BOLDRINI, J. L., COSTA, S. I. R., RIBEIRO, V. L. F. F., WETZLER, H. G., 1980, *Álgebra Linear*, 2ª Edição, São Paulo, Harper e Row do Brasil.

CAMPS-VALLS, G., PORTA-OLTRA, B., SORIA-OLIVAS, E., MARTIN-GUERRERO, J. D., SERRANO-LÓPEZ, A. J., PÉREZ-RUIXO, J. J., JIMÉNEZ-TORRES, N. V., 2003, “Prediction of Cyclosporine Dosage in Patients After Kidney Transplantation

Using Neural Networks”, *IEEE Transactions on Biomedical Engineering*, v. 50, n. 4, pp. 442-448.

CHANG, E., ABRAHAMOWICZ, M., FERLAND, D., FORTIN, P. R., 2002, “Comparison of the Responsiveness of Lupus Disease Activity Measures to Changes in Systemic Lupus Erythematosus Activity Relevant to Patients and Physicians”, *Journal of Clinical Epidemiology*, v. 55, pp. 488-497.

CLARK, T. G., ALTMAN, D. G., 2003, “Developing a Prognostic Model in The Presence of Missing Data: An Ovarian Cancer Case Study”, *Journal of Clinical Epidemiology*, v. 56, pp. 28-37.

COSTA, J.C.G.D., INFANTOSI, A. F. C., ALMEIDA, R. M. V., SUASSUNA, J. H., 2004, “Metodologia para a Determinação de Replicações “Bootstrap” para o Treinamento de Redes Neurais Artificiais: Uma Aplicação ao Transplante Renal”, In: *Proceedings of The International Federation for Medical and Biological Engineering*, João Pessoa, PB, Brasil, v.5, n.1, pp. 963-966, Sept 22-25.

CROSS, S. S., HARRISON, R. F., SANDERS, D. S., 2003, “Supporting Decisions in Clinical Medicine: Neural Networks in Lower Gastrointestinal Hemorrhage”, *The Lancet*, v. 362, n. 9392, pp. 1250-1251.

DAS, A., BEN-MENACHEN, T., COOPER, G. S., CHAK, A., SIVAK JR, M., V., GONET, J. A., WONG, R. C. K., 2003, “Prediction of Outcome in Acute Lower-Gastrointestinal Haemorrhage Based on an Artificial Neural Network: Internal and External Validation of a Predictive Model”, *The Lancet*, v.362, n. 9392, pp. 1261-1266.

DAVISON, A. C., HINKLEY, D. V., 1997, *Bootstrap Methods and Their Application*, Cambridge, Cambridge University Press.

de AZEVEDO, F. M., BRASIL, L. M., OLIVEIRA, R. C. L., 2000, *Redes Neurais com Aplicações em Controle e em Sistemas Especialistas*, Florianópolis, Visual Books, Bookstore.

DEMUTH, H., BEALE, M., 1998, *Neural Network Toolbox For Use With MATLAB – User's Guide*, Natick, MA, USA, The MathWorks, Inc

de SÁ, C. B. P., 1996, *Previsão de Séries Temporais Epidemiológicas Utilizando Redes Neurais*, Dissertação de Mestrado, Programa de Engenharia Biomédica, COPPE/UFRJ, Rio de Janeiro, 126 p., mar.

DORSEY, S. G., WALTZ, C. F., BROSCHE, L., CONNERNEY, I., SCHWEITZER, E. J., BARTLETT, S. T., 1997, "A Neural Network Model for Predicting Pancreas Transplant Outcome", *Diabetes Care*, v. 20, n. 7, pp. 1128-1133.

DVORCHIK, I., SUBOTIN, M., MARSH, W., MCMICHAEL, J., FUNG, J. J., 1996, "Performance of Multi-Layer Feedforward Neural Networks to Predict Liver Transplantation Outcome", *Methods of Information in Medicine*, v. 35, n. 1, pp. 12-18.

EFRON, B., 1994, "Missing Data, Imputation, and The Bootstrap", *Journal of the American Statistical Association*, v. 89, n. 426, pp. 463-475.

EFRON, B., 2000, "The Bootstrap and Modern Statistics", *Journal of the American Statistical Association*, v. 95, n. 452, pp. 1293-1296.

ENNETT, C. M., FRIZE, M., WALKER, C. R., 2001, "Influence of Missing Values on Artificial Neural Network Performance", In: *Proceedings of Medinfo, World Congress on Medical Informatics*, London, UK, pp. 449-453, Sept 2-5.

FISHER, N. I., HALL, P., 1991, "Bootstrap Algorithms for Small Samples", *Journal of Statistical Planning and Inference*, v. 27, pp. 157-169.

FURNESS, P. N., LEVESLEY, J., LUO, Z., TAUB, N., KAZI, J. I., BATES, W. D., NICHOLSON, M. L., 1999a, "A Neural Network Approach to the Biopsy Diagnosis of Early Acute Renal Transplant Rejection", *Histopathology*, v. 35, n. 5, pp. 461-467.

FURNESS, P. N., KAZI, J., LEVESLEY, J., TAUB, N., NICHOLSON, M., 1999b, "A Neural Network Approach to the Diagnosis of Early Acute Allograft Rejection", *Transplantation Proceedings*, v. 31, n. 8, pp. 3151.

GARCIA, V. D., GARCIA, C. D., KEITEL, E., ABBUD-FILHO, M., CAMPOS, H. H., PESTANA, J. O. M., 2003, "Living Kidney Transplantation in Brazil: Unwanted Procedure of Choice in View of Cadaver Organ Storage", *Transplantation Proceedings*, v. 35, n. 3, pp. 1182-1184.

GUARNIERI, G., SITULIN, R., BIOLO, G., 2001, "Carnitine Metabolism in Uremia", *American Journal of Kidney Diseases*, v. 38, n. 4 (suppl. 1), pp. S63-S67.

GISMONDI, R. C., 1999, *Modelagem da Mortalidade Infantil em Municípios Brasileiros por Redes Neurais e Regressão Linear Múltipla*, Tese de Doutorado, Programa de Engenharia Biomédica, COPPE/UFRJ, Rio de Janeiro, 182 p., mai.

GISMONDI, R. C., ALMEIDA, R. M. V., INFANTOSI, A. F. C., 2002, "Artificial Neural Network for Infant Mortality Modelling", *Computer Methods and Programs in Biomedicine*, v. 69, pp. 237-247.

GOLDFARB-RUMYANTZEV, A. S., SCANDLING, J. D., PAPPAS, L., SMOUT, R. J., HORN, S., 2003, "Prediction of 3-Year Cadaveric Graft Survival Based on Pre-Transplant Variables in a Large National Dataset", *Clinical Transplantation*, v.17, n. 6, pp. 485-497.

GUYTON, A. C., HALL, J. E., 1998, "Receptores Sensoriais; Circuitos Neurais Para o Processamento das Informações; Sentidos do Tato e da Posição", In: *Fisiologia Humana e Mecanismos das Doenças*, 6ª edição, Capítulo 32, Guanabara Koogan, Rio de Janeiro, Brasil.

HAHN, M. E., FARLEY, A. M., LIN, V., CHOU, L.C., 2005, "Neural Network Estimation of Balance Control During Locomotion", *Journal of Biomechanics*, v.38, pp.717-724.

HAYKIN, S., 2002, *Redes Neurais – Princípios e Prática*, 2ª Edição, Porto Alegre, Bookman.

HIRSH, R.P., RIEGELMAN, R.K., (1992), *Statistical First Aid*, Boston: Blackwell Scientific Publications.

HSIANG-YIN, C., TA-CHENG, C., MIN, D. I., FISCHER, G. W., WU, Y. M., 1999, "Prediction of Tacrolimus Blood Levels by Using the Neural Network with Genetic Algorithm in Liver Transplantation Patients", *Therapeutic Drug Monitoring*, v.21, n. 1, pp. 50-56.

HUGHES, V. F., MELVIN, D. G., NIRANJAN, M., ALEXANDER, G. A. M., TRULL, A. K., 2001, "Clinical Validation of an Artificial Neural Network Trained to Identify Acute Allograft Rejection in Liver Transplant Recipients", *Liver Transplantation*, v. 7, n. 6, pp. 496-503.

INFANTOSI, A. F. C., ALMEIDA, R. M. V. R., COSTA, J. C. G. D., SUASSUNA, J. H. R., 2004, "Definição de Variáveis em Redes Neurais Artificiais por Análise de Componentes Principais: Aplicação em Transplante Renal", In: *Anais do IX Congresso Brasileiro de Informática em Saúde*, v.1, pp.760-765, Ribeirão Preto, SP, Nov. 07-10.

JACOBS, R. A., 1988, "Increased Rates of Convergence Through Learning Rate Adaptation", *Neural Networks*, v.1, n.1, pp. 295-307.

JOHNSON, R., J., RUDGE, C., J., COLLETT, D., FORSYTHE, J., L., R., 2004, "Serum Creatinine In The First Year Post-Transplant As A predictor Of Long-term Renal Transplant Outcome In UK", In: *Proceedings Of American Transplant Congress, Poster Session: Kidney Transplant Outcomes And Predictors* (Tuesday, May 18, 2004, 12:30-2:30 PM), Boston, MA, USA, May 14-19.

JOHNSON, R. A., WICHERN, D. W., 1988, "Inferences About a Mean Vector", In: *Applied Multivariate Statistical Analysis*, 2<sup>nd</sup> Edition, Chapter 5, Englewood Cliffs, NJ, Prentice-Hall, USA.

KOVÁCS, Z. L., 2002, *Redes Neurais Artificiais – Fundamentos e Aplicações*, 3<sup>a</sup> Edição, Editora Livraria da Física, São Paulo

LABVIEW®, 2000, National Instruments Corporation, versão 6i.

LEITHOLD, L., 1977, “Cálculo Diferencial de Funções de Várias Variáveis”, In.: *O Cálculo com Geometria Analítica*, 2ª Edição, Volume 2, Capítulo 17, Editora Harper e Row do Brasil Ltda, São Paulo, Brasil.

LJUNG, L. 1999, *System Identification: Theory for the User*, 2<sup>nd</sup> Edition, Englewood Cliffs, NJ, Prentice-Hall.

LOPES A. A., BATISTA, P. B. P., COSTA F. A., NERY, M. M., LOPES, G. B., 2003, “Número de Anos em Tratamento Dialítico Crônico e Risco de Morte em Pacientes com e sem Diabetes Melito”, *Revista da Associação Médica Brasileira*, n. 49, v. 3, pp. 266-269.

LOONEY, S. W., 1995, “How to Use tests for Univariate Normality to Assess Multivariate Normality”, *The American Statistician*, v. 49, n. 1, pp. 64-70.

MAGALHÃES, R. A., SANCHES, M. D., PEREIRA, W. D., 1996, “O Doador”, In: *Manual de Transplantes de Órgãos e Tecidos*, Editor Walter A. Pereira, MEDSI Editora Médica e Científica Ltda, Capítulo 8, Rio de Janeiro, Brasil.

MANLY, B. F. J., 1997, *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2<sup>nd</sup> Edition, Eds. Chris Chatfield and Jim V. Zidek, Boundary Row, London, Chapman and Hall Texts in Statistical Science Series.

MANRIQUE, J., ERRASTI, P., LAVILLA, J., ROSSICH, E., HERNANDEZ, A., PUJANTE, D., GARCIA-FERNANDEZ, N., PURROY, A., TREATMENT OF HYPERHOMOCYSTEINEMIA AFTER RENAL TRANSPLANTATION”, *TRANSPLANTATION PROCEEDINGS*, V. 35, PP. 1742-1744.

MARSDEN, P. A., 2003, “Predicting Outcomes After Renal Transplantation – New Tools and Old Tools”, *New England Journal of Medicine*, v. 349, n. 2, pp. 182-184.

MATSUI, Y., SILVA-FILHO, A. P., COSTA, S. A. V., SILVA-JUNIOR, H. T., FRANCO, M. F., MEDEIROS, R., PESTANA, J. O. M., 2001, “Fatores Relacionados à Progressão da Disfunção Crônica do Enxerto em Transplantados Renais”, *Jornal Brasileiro de Nefrologia*, v. 23, n. 2, pp. 97-103.

MELVIN, D. G., NIRANJAN, M., PRAGER, R. W., TRULL, A. K., HUGHES, V. F., 2000, "Neuro-Computing Versus Linear Statistical Techniques Applied to Liver Transplant Monitoring: A Comparative Study", *IEEE Transactions on Biomedical Engineering*, v. 47, n. 8, pp. 1036-1043.

MENG, X. L., 2000, "Missing data: Dial M For ???", *Journal of the American Statistical Association*, v. 95, n. 452, pp. 1325-1330.

OOSTENBRINK, R., MOONS, K. G. M., BLEEKER, S. E., MOLL, H. A., GROBBEE, D. E., 2003, "Diagnostic Research on Routine Care Data: Prospects and Problems", *Journal of Clinical Epidemiology*, v. 56, pp. 501-506.

PARK, C., LINDSAY, B. G., 1999, "Robust Scale Estimation and Hypotesis Testing Based on Quadractic Inference Function", Technical Report, Nº 99-03, Center For Likelihood Studies, Department of Statistics, The Pennsylvania State University, University Park, PA, pp. 1-26

PEDHAZUR, E. J., 1982, "Regression Analysis With Categorical and Continuous Variables", In: *Multiple Regression in Behavioral Research*, 2<sup>nd</sup> Edition, Part 2, Eds. Holt, Rinehart and Winston, The Dryden Press, Saunders College Publishing, New York, USA.

PERCOITS-FILHO, R., STEVINKEL, P., LINDHOLM, B., BERGSTRÖN, J., NORONHA, I., ABENSUR, H., 2002, "Revisão: Desnutrição, Inflamação e Atherosclerose (Síndrome MIA) em Pacientes Portadores de Insuficiência Renal Crônica", *Jornal Brasileiro de Nefrologia*, v. 24, n. 3, pp. 136-146.

PERKOWSKA, A., ETHASADE, A., GACIONG, Z., DURLIK, M., PLACHA, G., GALAZKA, Z., LAO, M., 2002, "Regulators of Fibrinolysis in Plasma of Kidney Transplant Recipients", *Transplantation Proceedings*, v. 34, pp. 585-588.

POLITIS, D. N., ROMANO, J. P., 1994, "The Stationary Bootstrap", *Journal of the American Statistical Association*, v. 89, n.428, pp. 1303-1313.

QIAN, N., 1999, "On The Momentum Term in Gradient Descent Learning Algorithms", *Neural Networks*, v. 12, pp.145-151.

R: Copyright 2003, The R Development Core Team, Version 2.0.0, (2004), [www.r-project.org](http://www.r-project.org).

RADERMACHER, J., MENGEL, M., ELLIS, S., STUHT, S., HISS, M., SCHARZ, A., EISENBERGER, U., BURG, M., LUFT, F. C., GWINNER, W., HALLER, H., 2003, "The Arterial Resistance Index And Renal Allograft Survival", *The New England Journal Of Medicine*, v. 349, n.2, pp. 115-152.

REED, R., 1993, "Pruning Algorithms – A Survey", *IEEE Transactions on Neural Networks*, v. 4, n. 5, pp. 740-747.

RIELA, M. C., MARTINS, C., 2001, "Nutrição e Hemodiálise", In: *Nutrição e o Rim*, Eds. Miguel C. Riella e Cristina Martins, Capítulo 12, Guanabara Koogan, Rio de Janeiro, Brasil.

ROBINSON, K., 2004, "Renal Disease, Homocysteine, and Cardiovascular Complications", *Circulation*, v. 104, pp.294-295.

SALOMÃO-FILHO, A., SALAZAR, H. M., CÂMARA, F. P., RAMBALDUCCI, S., 1996, "Transplante Renal", In: *Manual de Transplantes de Órgãos e Tecidos*, Editor Walter A. Pereira, MEDSI Editora Médica e Científica Ltda, Capítulo 10, Rio de Janeiro, Brasil.

SARTORI, M., PATRASSI, G., RIGOTO, P., MARCHINI, F., FIORETTI, M., SPIEZIA, L., GIROLAMI, A., 2000, "Improved Fibrinolytic Capacity After Withdrawal of Steroid Immunosuppression in Renal Transplant Recipients", *Clinical Transplantation*, v. 69, n. 10, pp. 2116-2121.

SHEPPARD, D., MCPHEE, D., DARKR, C., SHRETHRA, B., MOORE, R., JUREWITZ, A., GRAY, A., 1999, "Predicting Citomegalovirus Disease After Renal Transplantation: An Artificial Network Approach", *International Journal of Medical Informatics*, v. 54, n. 1, pp. 55-76.

SHOSKES, D. A., TY, R., BARBA, L., SENDER, M., 1998, "Prediction of Early Graft Function in Renal Transplantation Using a Computer Neural Network", *Transplantation Proceedings*, v. 30, n. 4, pp. 1316-1317.

SIMIC-OGRIKOVIC, S., FURUNCIC, D., LEZAIK, V., RADIVIJEVIC, D., BLAGOJEVIC, R., DJUKANOVIC, L. J., 1999, "Using ANN in Selection of the Most Important Variables in Prediction of Chronic Renal Allograft Rejection Progression", *Transplantation Proceedings*, v. 31, n. 1-2, pp. 368.

SOLÁ, R., GUIRADO, L., DÍAZ, J.M., LÓPEZ-NAVIDAD, A., CABALLERO, F., GICH, I., (2002), "Elderly Donor Kidney Grafts Into Young recipients: Results at 5 Years", *Transplantation*, v. 73, n. 10, p. 1673-1681.

STEYERBERG, E. W., BLEEKER, S. E., HENRIËTTE, A. M., GROBBEE, D. E., MOONS, K. G. M., (2003), "Internal and External Validation of predictive Models: A Simulation Study of Bias and Precision in Small Samples", *Journal of Clinical Epidemiology*, v.56, pp. 441-447.

TAJINE, M., ELIZONDO, D., 1998, "The Recursive Deterministic Perceptron Neural Network", *Neural Networks*, v. 11, n. 9, pp. 1571-1588.

TERRIN, N., SCHIMID, C. H., GRIFFITH, J. L., D'AGOSTINO, R. B., SELKER, H. P., 2003, "External Validity of Predictive Models: A Comparison of Logistic Regression, Classification Trees, and Neural Networks", *Journal of Clinical Epidemiology*, v. 56, pp. 721-729.

TWISK, J., de VENT, W., 2002, "Attrition in Longitudinal Studies: How to Deal With Missing Data", *Journal of Clinical Epidemiology*, v. 55, pp. 329-337.

UNITED NETWORK FOR ORGAN SHARING, disponível em [www.unos.org](http://www.unos.org). Acessado em dezembro de 2004.

VERPOOTEN, G. A., COOLS, F. J., VAN DER PLANKEN, M. G., BEDERT, L. C., CLAES, C., VAN GAAL, L. F., DE BROE, M. E., 1996, "Elevated Plasminogen Inhibitor

Activator Levels in Cyclosporin-Treated Renal Allograft Recipients”, *Nephrology Dialysis Transplant*, v. 11, n. 2, pp. 347-351.

YAM, J. Y. F., CHOW, T. W. S., 2000, “A Weight Initialization Method for Improving Training Speed in Feedforward Neural Network”, *Neurocomputing*, v. 30, pp. 219-232.

ZHANG, J., 1999, “Inferential Estimation of Polymer Quality Using Bootstrap Aggregated Neural Network”. *Neural Networks*, v. 12, pp. 927-938.

# APÊNDICE A

## Tópicos em Redes Neurais Artificiais

### A.1 - Introdução

Semelhantemente ao comportamento neuronal humano (em muito menor escala do que acontece na realidade, pois há muito mais ramificação das fibras nervosas, atingindo também um número muito maior de neurônios, ver GUYTON e HALL, 1998), a Rede Neural Artificial (RNA) tem processamento paralelo e distribuído, propiciando algumas vantagens (BILLINGS *et al.*, 1992; GISMONDI, 1999; de AZEVEDO *et al.*, 2000; HAYKIN, 2002), tais como:

- a) Robustez quanto à perda de informações, por exemplo, a falta de um determinado sinal de entrada prejudica a qualidade da informação de saída, mas não de forma catastrófica. Pode ser verificado uma analogia RNA-Rede Neuronal Natural (RNN) quando da ocorrência de um AVC (Acidente Vascular Cerebral) de pequena intensidade, onde somente uma parcela das funções é perdida, de acordo com a extensão da área afetada;
- b) Rapidez no processamento das informações;
- c) Adaptabilidade em relação ao ambiente em que a rede está inserida. Através das mudanças dos pesos, a “força” da conexão é alterada. Analogia à própria teoria Darwiniana, onde aqueles que conseguem melhor e mais rápida adaptação ao meio têm maior probabilidade de sobrevivência;
- d) Mapeamento tipo “Caixa-Preta” (ou Aproximador de Funções), onde somente são disponíveis os dados de entrada e saída. Através de “Regras de Aprendizagem”, os pesos são alterados, seja em função da diferença entre o sinal desejado (vindo do ambiente externo) e o sinal de saída da rede (abordagem por correção do erro), seja em relação à correlação entre o sinal de entrada e saída do neurônio (abordagem *Hebbiana*). Não há nenhum conhecimento *a priori* do processo a ser modelado, conforme explicado anteriormente;
- e) Mapeamento através de funções não-lineares.

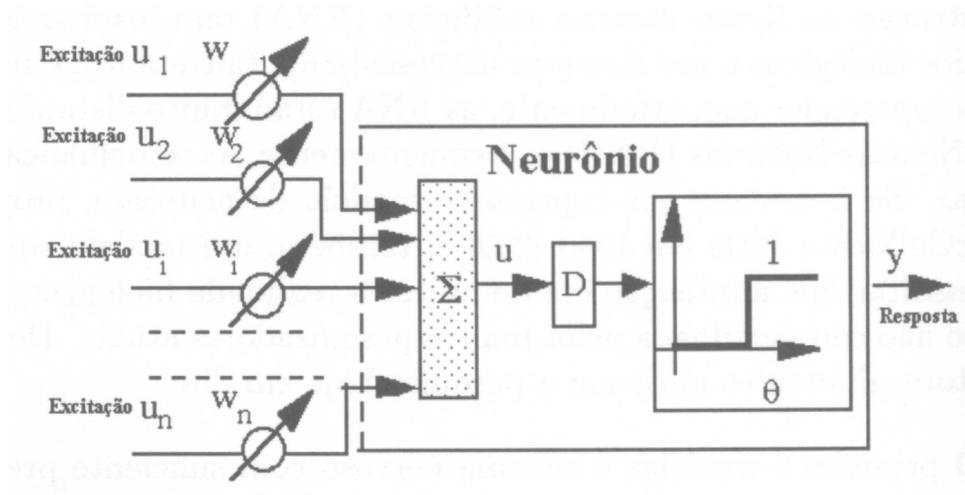
Embora a Rede Neural Artificial possa ser constituída de neurônios artificiais com função de ativação linear, a utilização de funções não-lineares é a mais comum,

possibilitando sua utilização na maioria dos casos reais, onde a não-linearidade é predominante.

A primeira aproximação de um neurônio foi descrita por *McCulloch e Pitts*, e é até hoje considerado o marco inicial no estudo das RNA (GISMONDI, 1999; de AZEVEDO *et al.*, 2000; HAYKIN, 2002). Seu modelo comportava um mapeamento do sinal de saída através de uma função linear do tipo tudo ou nada (função Limiar ou de *Heaviside*) aplicada a uma combinação linear de entradas binárias do sistema com seus coeficientes sendo os pesos. Caso a saída do combinador linear atingisse determinado limiar, o neurônio teria sua saída ativada. A função de *Heaviside*,  $h(u)$ , é dada por:

$$h(u) = \begin{cases} 1, & \text{se } u \geq 0 \\ 0, & \text{caso contrário} \end{cases} \quad (\text{A.1})$$

Abaixo, vemos o modelo de McCulloch-Pitts.



**Figura A.1** – Modelo de McCulloch e Pitts. *Extraído de AZEVEDO et al., 2000.*

Neste modelo, de AZEVEDO e colaboradores (2000) acrescentam que a saída do combinador linear pode sofrer um atraso de tempo (D), correspondendo ao tempo de difusão entre as sinapses biológicas. Normalmente esse termo é desprezado e a saída do combinador linear serve diretamente de argumento à função Limiar de saída.

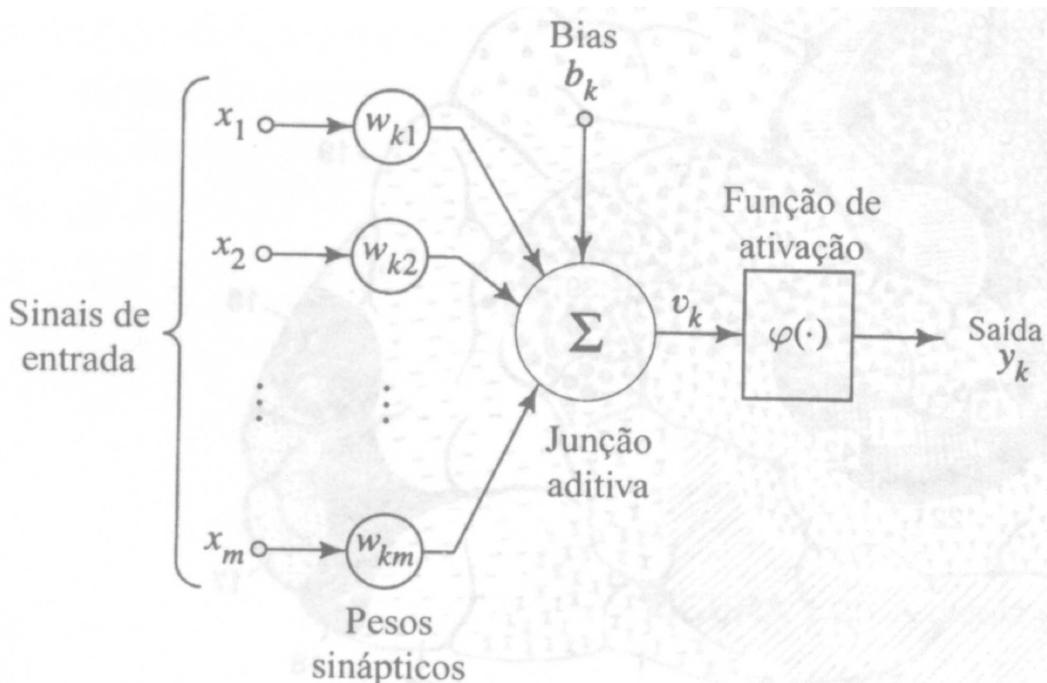
Para um aprofundamento no histórico do desenvolvimento das Redes Neurais Artificiais, ver GISMONDI (1999), de AZEVEDO (2000) e HAYKIN (2002).

Devido às suas características, tem havido interesse na identificação de modelos “Caixa-Preta” não-lineares de sistemas biológicos, onde a acessibilidade a certos sinais torna-se impraticável. As Redes Neurais Artificiais, por suas características de aprendizagem dinâmica, mimetizando as redes neuronais naturais (o conjunto de neurônios humanos, por exemplo) podem ser utilizadas para a identificação e predição de sistemas não-lineares deste tipo, pois são consideradas modelos “Aproximadores de Funções Universais” (HAYKIN, 2002). Em pacientes submetidos a transplante (TX), as decisões de tratamento são feitas com base no histórico do paciente, em exames clínicos e laboratoriais ou ainda, histopatológicos. Durante o período inicial, onde há chances de perda do enxerto devido a complicações clínicas ou cirúrgicas, o médico assistente avalia as variáveis envolvidas e implementa as ações necessárias. O comportamento da Rede Neural Artificial, baseado em treinamento de situações semelhantes, assemelha-se em grande parte ao mesmo comportamento do médico assistente, o que pode ser de grande auxílio diagnóstico (DORSEY *et al.*, 1997).

As Redes Neurais Artificiais podem ser classificadas de acordo com sua topologia. Redes Alimentadas Diretamente (*Feedforward*), Uni ou Multi-Camadas ou ainda Redes Recorrentes (com realimentação), Uni ou Multi-Camadas (BILLINGS *et al.*, 1992; de AZEVEDO *et al.*, 2000; HAYKIN, 2002). Dependendo do número de saídas, são mais comumente enquadradas como MISO (*Multiple-Input, Single-Output*) ou MIMO (*Multiple-Input, Multiple-Output*). Suas funções de ativação podem ser lineares ou não-lineares, sendo as mais comuns as de formato sigmoidal, tais como a função Logística e a função Tangente Hiperbólica. A função mais utilizada é a Logística, dada por:

$$f(u) = \frac{1}{1 + e^{-\alpha \cdot u}} \quad (\text{A.2})$$

onde  $\alpha$  é um parâmetro que determina a inclinação da curva. Quando  $\alpha$  tende a infinito, a função Logística torna-se a função Limiar. É importante verificar que a função Logística é diferenciável e contínua no intervalo [0,1]. Exemplo de uma RNA com somente um neurônio com várias entradas e uma saída é mostrado na Figura A.2:



**Figura A.2** - Modelo de RNA com 1 neurônio. A saída é o resultado da aplicação da Função de Ativação (FA) à combinação linear ( $v_k$ ) das  $m$  entradas  $x_1 \dots x_m$  com os pesos correspondentes  $w_{k1} \dots w_{km}$ , mais o *bias* ( $b_k$ ). *Extraído de HAYKIN, 2002.*

## A.2- Regras de aprendizagem por correção do erro

As regras de aprendizagem influenciam na alteração dos valores dos pesos, e conseqüentemente da sua “força”. A aprendizagem por correção do erro é função do erro encontrado, a cada iteração, entre o sinal de saída do neurônio (ou da Rede Neural Artificial) e o valor desejado (do ambiente externo). A atualização do peso é feita somando-se seu valor inicial com a variação encontrada na iteração, que é proporcional ao produto do erro e um parâmetro definido como Taxa de Aprendizagem ( $\eta$ ). Na aprendizagem conhecida como *Supervisionada*, para que o erro seja determinado, é necessário que a rede tenha conhecimento do sistema, onde então o erro é dado pela diferença entre a saída da rede e a saída desejada,  $e(n) = d(n) - y(n)$ , onde  $e(n)$ ,  $d(n)$  e  $y(n)$  são o erro, o valor desejado e a saída da RNA na iteração  $n$ , respectivamente.

Usualmente o erro nesta abordagem é o Erro Quadrático Médio (EQM), e o EQM pode ser entendido como a energia instantânea da diferença entre a saída da Rede

Neural Artificial e o sistema externo. Este erro é realimentado para atualização de valores das conexões sinápticas até que seu valor seja minimizado. Os algoritmos que utilizam esta abordagem são a Regra Delta, ou de *Widrow-Hopf* e a Regra Delta Generalizada, ou algoritmo de Retropropagação (HAYKIN, 2002). O algoritmo de Retropropagação (do inglês *Back-Propagation - BP*), junto com suas variações, é um dos métodos mais utilizados no aprendizado das RNA. A atualização do peso  $w(n)$  na iteração  $n$ , na regra Delta, é dado por:

$$w(n) = w(n-1) + \eta \cdot e(n-1) \cdot x(n-1), \quad (\text{A.3})$$

onde  $w(n-1)$ ,  $x(n-1)$  e  $e(n-1)$  são o peso, a entrada e o erro na iteração  $n-1$ , respectivamente.

A taxa de aprendizagem  $\eta$  é de muita importância, à medida que seu valor pode levar à divergência do processo iterativo. SHEPPARD *et al.* (1999) utilizaram taxas de aprendizagem adaptativas, e sugeriram que as mesmas devam começar com valores entre 0,5 e 0,05 e diminuir à medida que a iteração avança. Tal método também foi utilizado por de SÁ (1996). Tais valores iniciais adaptativos (ou fixos) estão de acordo com o preconizado por HAYKIN (2002), DEMUTH e BEALE (1998) e ANDREA e KALAYEH (1991), de que as taxas de aprendizagem devem situar-se entre 0 e 1.

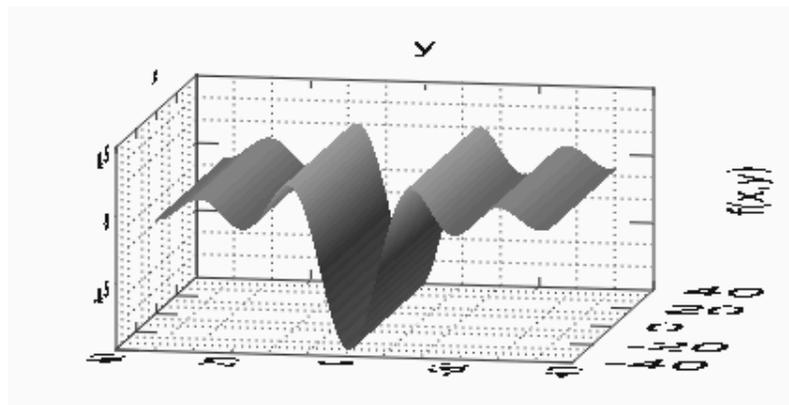
### A.3– Técnicas de otimização – Gradiente Descendente

“O sistema real é uma entidade esotérica que não pode ser alcançado por modelagem prática” (LJUNG, 1999). Esta afirmação traduz, na verdade, que sempre existirá uma diferença entre a saída do modelo e a do sistema real, em qualquer tipo de modelagem. E para que a diferença seja a menor possível, é necessário o estabelecimento de uma combinação de parâmetros ótimos do modelo, que são determinados a partir da minimização (ou maximização, dependendo do problema) da função objetivo (também chamada de função custo ou índice de desempenho) e do método de minimização / maximização da mesma. Logo, dependendo da função objetivo e do método de minimização / maximização, diferentes combinações de parâmetros podem ser encontradas. Como os métodos de minimização / maximização visam encontrar parâmetros ótimos, é comum chamá-los de métodos de otimização.

Em funções contínuas de duas variáveis, os pontos no  $\mathcal{R}^3$  que se deslocam ao longo de uma superfície o fazem segundo as relações entre as variáveis independentes e

dependente. Por exemplo, seja uma função  $f(x, y) = z = -\text{sinc}\left(\sqrt{\frac{x^2}{16} + \frac{y^2}{9}}\right)$ , cujo

gráfico está na Figura A.3:



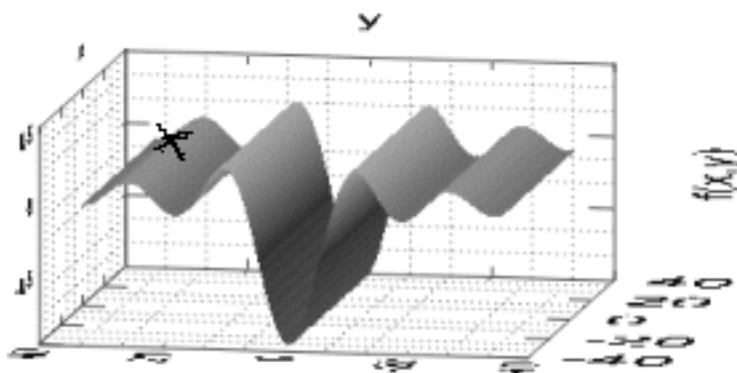
**Figura A.3** - Superfície  $f(x,y)$  no espaço vetorial  $\mathbb{R}^3$ , gerada em LabView®.

Da mesma forma que no  $\mathcal{R}^2$ , para se obter um ponto mínimo  $P(x_0, y_0, z_0)$  em três dimensões, seja ele, relativo ou absoluto, deriva-se a função em relação a cada variável independente e anula-se a expressão. Se a função  $f(x, y) = z$  for parametrizada em uma função  $\varphi(x, y, z)$ , e se derivarmos esta função em relação a cada variável, encontraremos um vetor perpendicular a um plano tangente em um ponto e este vetor é chamado de *gradiente* da função. O vetor  $\frac{\partial \varphi}{\partial x} \vec{i} + \frac{\partial \varphi}{\partial y} \vec{j} + \frac{\partial \varphi}{\partial z} \vec{k}$  é o gradiente de  $\varphi(x, y, z)$  e é representado por  $\nabla \varphi$ , onde  $\vec{i}$ ,  $\vec{j}$  e  $\vec{k}$  são os vetores unitários nas direções positivas de  $x$ ,  $y$  e  $z$  (LEITHOLD, 1977). Raciocínio análogo pode ser feito para funções de mais de três variáveis, embora a visualização não seja possível.

Os métodos de otimização baseados no gradiente da função objetivo em relação aos pesos visam a minimização da função na direção contrária ao do crescimento do vetor gradiente. Se o módulo do gradiente está aumentando, significa que a variação da função em relação aos pesos também está aumentando, ou seja, devemos ir na direção contrária, se desejarmos encontrar um mínimo local ou global. O método do

Gradiente Descendente (GD) “procura” a solução na direção do decréscimo do módulo do gradiente da função, e quanto maior for sua variação, maior será a alteração.

Para ilustração, é apresentado a seguir um exemplo simples, de funções de duas variáveis. Dependendo da posição inicial do ponto X (Figura A.4), ao percorrer um caminho à esquerda, este ponto nunca encontrará o mínimo global, enquanto que se percorrer em direção à direita poderá encontrá-lo, dependendo da sua capacidade de reconhecimento do mínimo local que está mais próximo.



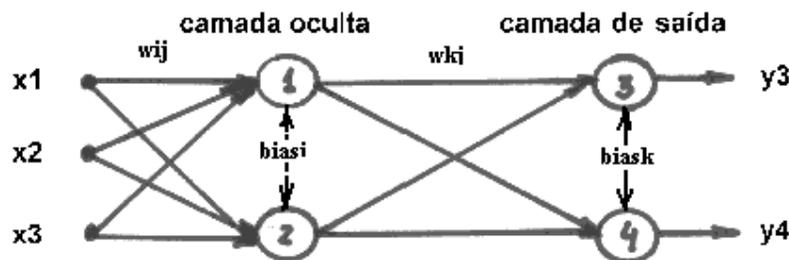
**Figura A.4** - Percurso do ponto X sobre a superfície de erro  $f(x,y)$ , simulada em LabVIEW®

A otimização pelo Gradiente Descendente não garante a obtenção do mínimo global (BILLINGS *et al.*, 1992; BABA, 1994; HAYKIN, 2002). Para que o algoritmo tenha a capacidade de convergir para um mínimo global, evitando os mínimos locais, alguns autores desenvolveram técnicas otimizadas para a escolha dos pontos iniciais (YAM e CHOW, 2000), ou ainda criaram variantes do GD, como o Gradiente Descendente Conjugado (GDC), combinação de GDC com função Gaussiana, ou o acréscimo do termo de Momento, ou ainda incluíram  $\eta$  adaptativo (BILLINGS *et al.*, 1992; BABA *et al.*, 1994; de SÁ, 1996; GISMONDI, 1999; de AZEVEDO, 2000; HAYKIN, 2002). O termo de Momento pode ser entendido como uma tentativa de estabilização do algoritmo na procura do mínimo global. Uma analogia interessante da constante de momento com um corpo em movimento em meio viscoso é introduzida por QIAN (1999). Sua expressão é proporcional à atualização do peso da iteração

anterior. O termo de Momento será mais detalhado posteriormente, após a apresentação do algoritmo BP.

#### A.4 – Algoritmo de treinamento por Retropropagação de Erro (BP)

O algoritmo BP é o utilizado neste trabalho, e, para sua apresentação, será utilizada uma Rede Neural Artificial com 3 entradas, 2 neurônios ocultos na camada  $j$  e 2 neurônios de saída na camada  $k$ , uma arquitetura 3x2x2 (Figura A.5). Considerando as funções de ativação diferenciáveis e contínuas, e iguais, mais a constante *bias* para todos os neurônios e o método de otimização o Gradiente Descendente, tem-se que a entrada da rede é o vetor  $\mathbf{X}=(x_1, x_2, x_3)$ , os pesos são  $w_{ij}$  e  $w_{kj}$ , onde  $j$  é o índice das entradas,  $i$  é o índice dos neurônios da camada oculta e  $k$  é o índice dos neurônios da camada de saída. As saídas são  $y_k$ .



**Figura A.5** - Modelo de rede alimentada diretamente com três entradas, uma camada de entrada, uma camada oculta  $i$  (neurônios 1 e 2) e uma camada de saída  $k$  (neurônios 3 e 4). As funções de ativação são as mesmas para todos os neurônios.

Pode-se então equacionar os argumentos  $v_i$  e  $v_k$  das funções de ativação  $\varphi(v_i)$  e  $\varphi(v_k)$ , respectivamente.

$$v_1 = w_{11} \cdot x_1 + w_{12} \cdot x_2 + w_{13} \cdot x_3 + bias_1$$

$$\therefore v_1 = \sum_{j=1}^3 w_{ij} \cdot x_j + bias_1$$

(A.4)

Similarmente, para  $v_2$ , tem-se:

$$v_2 = \sum_{j=1}^3 w_{2j} \cdot x_j + \text{bias}_2 \quad (\text{A.5})$$

e, para  $v_3$  e  $v_4$ :

$$v_3 = w_{31} \cdot y_1 + w_{32} \cdot y_2 + \text{bias}_3 \quad (\text{A.6})$$

$$\therefore v_3 = \sum_{k=1}^2 w_{3k} \cdot y_k + \text{bias}_3$$

$$v_4 = \sum w_{4k} \cdot y_k + \text{bias}_4 \quad (\text{A.7})$$

Para as saídas dos neurônios ocultos 1 e 2:

$$y_1 = \varphi(v_1) \quad (\text{A.8})$$

$$y_2 = \varphi(v_2) \quad (\text{A.9})$$

e para os neurônios de saída 3 e 4:

$$y_3 = \varphi(v_3) \quad (\text{A.10})$$

$$y_4 = \varphi(v_4) \quad (\text{A.11})$$

onde o erro da saída ( $e_k$ ) de cada neurônio da camada  $k$  é dado por:

$$e_3 = (d - y_3) \quad (\text{A.12})$$

$$e_4 = (d - y_4) \quad (\text{A.13})$$

e  $d$  é a saída desejada do sistema. A função objetivo na camada  $k$  é o EQM e é dado por:

$$\text{EQM} = \frac{1}{2} \cdot [e_3^2 + e_4^2] = \frac{1}{2} \sum e_k^2 \quad (\text{A.14})$$

A variação do peso entre o neurônio de saída 3 e o de entrada 1, e utilizando a regra da cadeia do cálculo diferencial e integral, tem-se que:

$$\Delta w_{31} = -\eta \cdot \frac{\partial \text{EQM}}{\partial w_{31}} = -\eta \cdot \frac{\partial \text{EQM}}{\partial v_3} \cdot \frac{\partial v_3}{\partial w_{31}} \quad (\text{A.15})$$

Mas:

$$\begin{aligned}
\frac{\partial EQM}{\partial v_3} &= \frac{\partial[(d - y_3)^2 + (d - y_4)^2]}{2 \cdot \partial v_3} = \frac{1}{2} \cdot \frac{\partial[(d - \varphi(v_3))^2]}{\partial v_3} = \\
&= \frac{1}{2} \cdot (-2) \cdot \varphi'(v_3) \cdot (d - y_3) = \\
&= -\varphi'(v_3) \cdot (d - y_3) = \\
&= -\delta_3
\end{aligned} \tag{A.16}$$

uma vez que  $\varphi(v_3)$  é diferenciável e contínua. Continuando...

$$\begin{aligned}
\frac{\partial v_3}{\partial w_{31}} &= \frac{\partial}{\partial w_{31}} [w_{31} \cdot y_1 + w_{32} \cdot y_2 + bias_3] = \\
&= y_1.
\end{aligned} \tag{A.17}$$

então:

$$\begin{aligned}
\Delta w_{31} &= -\eta \left[ -\varphi'(v_3) \cdot (d - y_3) \right] \cdot y_1 = \\
&= \eta \cdot \delta_3 \cdot y_1
\end{aligned} \tag{A.18}$$

Similarmente,

$$\begin{aligned}
\Delta w_{32} &= -\eta \left[ -\varphi'(v_3) \cdot (d - y_3) \right] \cdot y_2 = \\
&= \eta \cdot \delta_3 \cdot y_2
\end{aligned} \tag{A.19}$$

e, para o neurônio 4:

$$\begin{aligned}
\Delta w_{41} &= -\eta \left[ -\varphi'(v_4) \cdot (d - y_4) \right] \cdot y_1 = \\
&= \eta \cdot \delta_4 \cdot y_1 \\
\Delta w_{42} &= -\eta \left[ -\varphi'(v_4) \cdot (d - y_4) \right] \cdot y_2 = \\
&= \eta \cdot \delta_4 \cdot y_2
\end{aligned} \tag{A.20}$$

onde

$$\varphi'(v_4) \cdot (d - y_4) = \delta_4 = -\frac{\partial EQM}{\partial v_4} \tag{A.21}$$

O problema agora é achar a variação dos pesos referentes às conexões entre as entradas da RNA e os neurônios ocultos. A determinação destes pesos é feita propagando-se para trás (“retropropagando”, daí o nome do algoritmo) o erro da

camada de saída para a camada oculta, levando-se em consideração a contribuição de cada neurônio de saída. Utilizando-se a regra da cadeia novamente:

$$\begin{aligned} \Delta w_{11} \Big|_{\text{contrib. do neurônio3}} &= \\ &= -\eta \cdot \frac{\partial EQM}{\partial v_3} \cdot \frac{\partial v_3}{\partial y_1} \cdot \frac{\partial y_1}{\partial v_1} \cdot \frac{\partial v_1}{\partial w_{11}} \end{aligned} \quad (A.22)$$

$$\begin{aligned} \Delta w_{11} \Big|_{\text{contrib. do neurônio4}} &= \\ &= -\eta \cdot \frac{\partial EQM}{\partial v_4} \cdot \frac{\partial v_4}{\partial y_1} \cdot \frac{\partial y_1}{\partial v_1} \cdot \frac{\partial v_1}{\partial w_{11}} \end{aligned} \quad (A.23)$$

Calculando cada uma das parcelas:

$$\begin{aligned} \frac{\partial v_3}{\partial y_1} &= \frac{\partial}{\partial y_1} [w_{31} \cdot y_1 + w_{32} \cdot y_2 + \text{bias}_3] = \\ &= w_{31} \end{aligned} \quad (A.24)$$

$$\begin{aligned} \frac{\partial y_1}{\partial v_1} &= \frac{\partial}{\partial v_1} [\varphi(v_1)] = \\ &= \varphi'(v_1) \end{aligned} \quad (A.25)$$

$$\begin{aligned} \frac{\partial v_1}{\partial w_{11}} &= \frac{\partial}{\partial w_{11}} [w_{11} \cdot x_1 + w_{12} \cdot x_2 + w_{13} \cdot x_3 + \text{bias}_1] = \\ &= x_1 \end{aligned} \quad (A.26)$$

$$\begin{aligned} \frac{\partial v_4}{\partial y_1} &= \frac{\partial}{\partial y_1} [w_{41} \cdot y_1 + w_{42} \cdot y_2 + \text{bias}_4] = \\ &= w_{41} \end{aligned} \quad (A.27)$$

Utilizando-se as equações (A.24) e (A.29), e sabendo-se que:

$$\Delta w_{11} = \Delta w_{11} \Big|_{\text{contrib. do neurônio3}} + \Delta w_{11} \Big|_{\text{contrib. do neurônio4}} \quad (A.28)$$

$\Delta w_{11}$  torna-se:

$$\begin{aligned}
\Delta w_{11} &= -\eta \cdot w_{31} \cdot \varphi'(v_1) \cdot x_1 \cdot \left[ -\varphi'(v_3) \cdot (d - y_3) \right] - \\
&- \eta \cdot w_{41} \cdot \varphi'(v_1) \cdot x_1 \cdot \left[ -\varphi'(v_4) \cdot (d - y_4) \right] = \\
&= \eta \cdot \varphi'(v_1) \cdot x_1 \cdot \left[ \delta_3 \cdot w_{31} + \delta_4 \cdot w_{41} \right]
\end{aligned} \tag{A.29}$$

Similarmente,

$$\begin{aligned}
\Delta w_{12} &= \eta \cdot \varphi'(v_1) \cdot x_2 \cdot \left[ \delta_3 \cdot w_{31} + \delta_4 \cdot w_{41} \right] \\
\Delta w_{13} &= \eta \cdot \varphi'(v_1) \cdot x_3 \cdot \left[ \delta_3 \cdot w_{31} + \delta_4 \cdot w_{41} \right]
\end{aligned} \tag{A.30}$$

e

$$\begin{aligned}
\Delta w_{21} &= \eta \cdot \varphi'(v_2) \cdot x_1 \cdot \left[ \delta_3 \cdot w_{32} + \delta_4 \cdot w_{41} \right] \\
\Delta w_{22} &= \eta \cdot \varphi'(v_2) \cdot x_2 \cdot \left[ \delta_3 \cdot w_{32} + \delta_4 \cdot w_{42} \right] \\
\Delta w_{23} &= \eta \cdot \varphi'(v_2) \cdot x_3 \cdot \left[ \delta_3 \cdot w_{32} + \delta_4 \cdot w_{42} \right]
\end{aligned} \tag{A.31}$$

Generalizando para  $m$  neurônios na camada de saída, a variação dos pesos na camada de entrada fica:

$$\Delta w_{ij} = \eta \cdot \varphi'(v_i) \cdot x_j \cdot \sum_{k=1}^m \delta_k \cdot w_{ki} \tag{A.32}$$

ou

$$\Delta w_{ij} = \eta \cdot \delta_i \cdot x_j \tag{A.33}$$

onde

$$\delta_i = \varphi'(v_i) \cdot \sum_{k=1}^m \delta_k \cdot w_{ki} \tag{A.34}$$

e na camada de saída:

$$\Delta w_{ki} = \eta \cdot \delta_k \cdot y_i \tag{A.35}$$

A atualização dos pesos na camada oculta fica definida como:

$$w_{ij}(n) = w_{ij}(n-1) + \Delta w_{ij}(n-1) \tag{A.36}$$

e na camada de saída:

$$w_{ki}(n) = w_{ki}(n-1) + \Delta w_{ki}(n-1) \tag{A.37}$$

Incluindo os termos de Momento, tem-se na camada oculta:

$$w_{ij}(n) = w_{ij}(n-1) + \Delta w_{ij}(n-1) + \alpha \Delta w_{ij}(n-2) \quad (\text{A.38})$$

onde a constante  $\alpha$  é chamada de constante de momento. Na camada de saída:

$$w_{ki}(n) = w_{ki}(n-1) + \Delta w_{ki}(n-1) + \alpha \Delta w_{ki}(n-2) \quad (\text{A.39})$$

HAYKIN (2002) descreve quatro propriedades principais do uso do termo de Momento:

- a) Para que  $\Delta w$  seja convergente,  $|\alpha|$  deve situar-se entre 0 e 1;
- b) Se em iterações sucessivas, o sinal de  $\frac{\partial EQM}{\partial w_{ij,ki}}$  não varia,

significa que o algoritmo está em uma declividade decrescente. Neste caso, o termo de momento acelera a descida;

- c) No caso contrário ao do item b), se o sinal varia de uma iteração para outra, o termo de momento atenua esta variação;
- d) O uso do termo de momento pode evitar a convergência do algoritmo em um mínimo local raso.

O algoritmo BP, ao “retropropagar” o erro dos neurônios de saída para as camadas anteriores, estabelece uma contribuição de cada peso para o erro da saída da RNA. Simplificadamente, o algoritmo BP funciona como se segue:

- a) Comparação entre as saídas da RNA e o valor desejado dado pelo sistema;
- b) O erro encontrado é retropropagado para as camadas anteriores;
- c) Ajuste dos pesos utilizando o método de otimização GD;
- d) Cálculo das novas saídas da RNA;
- e) Comparação entre as novas saídas e o valor desejado do sistema;
- f) Repete-se o ciclo até a convergência do erro.

## APÊNDICE B

### Características das Variáveis

#### B.1 – Relação dos pacientes (receptores)

**Tabela B.1** - Relação dos 109 receptores seleccionados, de Janeiro de 1986 até Dezembro de 1995.

rec	data tx	Sobr	rec	data tx	Sobr	rec	data tx	Sobr	Rec	data tx	Sobr
#1	30/01/86	108,6	#31	25/06/87	97,4	#61	12/1/88	46,4	#91	24/01/91	48,7
#2	22/05/86	47,7	#32	7/2/87	97,4	#62	15/12/88	73,0	#92	4/4/91	48,7
#3	7/3/86	4,2	#33	7/9/87	97,4	#63	3/3/88	85,2	#93	25/04/91	31,3
#4	27/11/86	98,6	#34	23/07/87	18,4	#64	21/04/88	85,2	#94	7/10/91	48,7
#5	23/01/86	30,1	#35	17/09/87	61,9	#65	30/06/88	67,0	#95	23/07/91	48,7
#6	20/02/86	0,1	#36	24/09/87	52,0	#66	17/11/88	86,7	#96	23/07/91	51,8
#7	20/03/86	46,1	#37	10/8/87	85,2	#67	24/11/88	73,0	#97	15/08/91	48,7
#8	4/3/86	2,4	#38	15/10/87	97,4	#68	12/1/89	73,0	#98	10/3/91	48,7
#9	17/04/86	98,2	#39	22/10/87	11,5	#69	2/9/89	73,0	#99	31/10/91	36,5
#10	5/8/86	66,9	#40	29/10/87	146,1	#70	16/02/89	60,9	#100	13/11/91	36,1
#11	19/06/86	109,6	#41	19/11/87	97,4	#71	7/6/89	73,0	#101	13/2/92	0,0
#12	21/08/86	8,5	#42	17/03/88	48,7	#72	8/10/89	60,2	#102	12/11/92	24,3
#13	9/3/86	66,2	#43	24/03/88	73,0	#73	28/09/89	11,7	#103	26/11/92	24,3
#14	9/11/86	109,6	#44	4/4/88	5,0	#74	19/10/89	73,0	#104	3/3/94	12,2
#15	10/2/86	109,6	#45	14/04/88	43,9	#75	18/01/90	0,0	#105	20/04/94	0,0
#16	11/6/86	4,9	#46	28/04/88	73,0	#76	25/01/90	60,9	#106	19/05/94	12,2
#17	13/11/86	48,2	#47	5/5/88	26,2	#77	16/03/90	60,9	#107	29/09/94	12,2
#18	16/01/86	109,6	#48	5/12/88	85,2	#78	29/03/90	78,3	#108	17/11/94	12,2
#19	27/02/86	0,0	#49	19/05/88	19,7	#79	17/05/90	21,0	#109	15/12/94	12,2
#20	13/03/86	109,6	#50	26/05/88	60,9	#80	24/05/90	60,9			
#21	28/08/86	0,0	#51	6/2/88	79,2	#81	21/06/90	5,4			
#22	13/08/87	73,6	#52	23/06/88	85,2	#82	26/07/90	24,4			
#23	3/12/87	97,4	#53	7/7/88	85,2	#83	8/9/90	45,4			
#24	19/03/87	97,4	#54	22/07/88	0,1	#84	13/09/90	23,8			
#25	30/03/87	5,6	#55	8/4/88	85,2	#85	12/6/90	32,2			
#26	14/05/87	51,3	#56	8/11/88	0,7	#86	13/12/90	26,0			

#27	18/05/87	97,4	#57	9/1/88	31,6	#87	20/12/90	0,0
#28	28/05/87	36,8	#58	9/8/88	13,2	#88	5/1/90	60,9
#29	6/4/87	102,3	#59	15/09/88	40,1	#89	15/01/90	1,6
#30	6/11/87	97,4	#60	10/6/88	85,2	#90	29/11/90	48,7

## B.2 – Variáveis dependente e preditoras

**Tabela B.2** – Variáveis preditoras originais. As unidades das variáveis contínuas estão entre parênteses e as categorias das variáveis dicotômicas e categóricas estão discriminadas na coluna tipo de variáveis. Os maiores valores das codificações das variáveis dicotômicas (valor 1) e categóricas (código 3) são aqueles que contribuem para a longevidade do enxerto. *Extraído da Tabela IV.1.*

Variáveis	Símbolo	Tipo da Variável
<i>Idade Receptor</i> (anos)	E1	Contínua
<i>Idade Doador</i> (anos)	E2	Contínua
<i>Sexo Receptor</i>	E3	Dicotômica, f=1
<i>Sexo Doador</i>	E4	Dicotômica, f=0
<i>Raça Receptor</i>	E5	Dicotômica, afro-descen=0
<i>Raça Doador</i>	E6	Dicotômica, afro-descen=0
<i>Transplantes Prévios</i>	E7	Numérica
<i>Transfusões Sangüíneas Receptor</i>	E8	Numérica
<i>Tempo de Diálise</i> (meses)	E9	Numérica
<i>Tipo da Diálise</i> (HD, CAPD)	E10	Dicotômica, HD =0
<i>Rejeição</i>	E11	Dicotômica, não = 1
<i>Compatibilidade HLA</i> (Doador Vivo)	E12	Categ, Id=3, Haplo=2, Não-Relacionado=1
<i>Esquema Imunossupressor</i>	E13	Dicotômica, Uso de CsA=1

**Tabela B.3** – Vetores das variáveis preditoras originais, por receptor. As unidades e codificação das variáveis estão mostradas na Tabela B.2 e no Capítulo IV.

Receptor	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13
#1	38	25	1	0	0	1	0	4	13,1	0	1	1	1
#2	31	29	1	1	1	1	0	14	19,9	0	1	1	0
#3	31	24	0	0	1	1	0	10	14,3	1	1	1	0
#4	39	37	0	0	0	0	0	5	21,2	0	1	1	0

**Tabela B.3 - Cont.**

<b>Receptor</b>	<b>E1</b>	<b>E2</b>	<b>E3</b>	<b>E4</b>	<b>E5</b>	<b>E6</b>	<b>E7</b>	<b>E8</b>	<b>E9</b>	<b>E10</b>	<b>E11</b>	<b>E12</b>	<b>E13</b>
#5	16	20	1	1	1	1	0	26	17,0	0	1	2	1
#6	41	35	1	1	0	0	0	3	33,2	0	1	2	0
#7	35	59	1	0	0	0	0	4	12,8	0	1	2	0
#8	34	38	1	1	1	1	0	6	5,1	0	1	2	1
#9	34	67	1	0	0	0	0	8	25,9	0	1	2	0
#10	9	36	1	1	1	1	0	9	13,3	0	1	2	1
#11	33	57	1	1	1	1	0	8	12,8	0	1	2	1
#12	52	55	1	1	1	1	0	8	51,4	0	1	2	1
#13	10	50	0	1	0	0	0	5	10,4	0	1	2	1
#14	41	47	1	0	1	1	0	5	9,4	0	1	2	1
#15	27	25	1	0	1	1	0	6	5,4	0	1	2	1
#16	53	34	0	0	1	1	0	8	40,9	0	1	2	1
#17	37	30	0	1	1	1	0	26	28,8	0	1	2	1
#18	33	31	0	1	0	0	0	10	38,1	0	1	3	0
#19	57	58	0	1	0	0	0	8	24,3	0	1	3	0
#20	55	48	0	0	0	0	0	13	15,6	0	1	3	0
#21	37	32	1	0	1	1	0	4	41,5	0	1	3	0
#22	48	44	0	0	1	1	0	8	25,8	0	1	3	0
#23	21	19	1	0	0	0	0	9	44,7	0	1	2	1
#24	41	39	1	1	1	1	0	21	19,8	0	1	2	1
#25	38	42	0	0	0	0	0	2	25,3	0	1	2	1
#26	23	49	1	0	0	0	0	4	17,6	0	1	2	1
#27	50	50	0	1	1	1	0	13	11,7	0	1	2	1
#28	10	46	0	1	0	0	0	7	18,1	0	1	2	1
#29	30	26	1	0	0	0	0	16	23,5	0	1	2	1
#30	38	30	1	0	1	1	0	5	17,4	0	1	2	1
#31	6	28	0	1	0	1	0	20	8,9	0	1	2	1
#32	34	22	1	1	0	1	0	2	29,6	0	1	2	1
#33	51	44	1	0	0	0	0	15	49,9	0	1	2	1
#34	38	40	1	0	0	0	0	2	20,0	0	1	2	1
#35	11	35	0	1	1	1	0	3	12,7	0	1	2	1

**Tabela B.3 - Cont.**

Receptor	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13
#36	16	44	0	0	1	1	0	14	3,8	0	1	2	1
#37	32	49	1	1	0	0	1	24	4,4	0	1	2	1
#38	14	39	0	1	0	0	0	9	33,9	0	1	2	1
#39	49	24	1	0	1	1	0	14	12,9	0	1	2	1
#40	23	48	0	1	1	1	0	2	35,4	0	1	2	1
#41	9	30	0	1	1	1	0	4	9,7	0	1	2	1
#42	23	31	0	1	0	0	0	7	25,8	0	1	2	1
#43	16	47	0	1	0	0	0	10	18,0	0	1	2	1
#44	13	35	0	1	0	0	0	1	3,1	0	1	2	1
#45	26	53	1	0	1	1	0	9	13,7	0	1	2	1
#46	24	27	1	1	0	0	0	15	27,2	0	1	2	1
#47	50	25	0	1	0	0	0	2	22,5	0	1	2	1
#48	37	29	0	1	1	1	0	21	26,5	0	1	2	1
#49	21	50	1	1	0	0	0	5	50,3	0	1	2	1
#50	46	43	1	0	0	0	0	4	28,2	0	1	2	1
#51	25	47	1	1	0	0	0	3	9,4	0	1	2	1
#52	32	31	1	1	1	1	0	6	32,2	0	1	2	1
#53	39	26	1	1	1	1	0	22	27,6	0	1	2	1
#54	64	60	1	1	1	1	0	3	18,9	0	1	2	1
#55	34	35	1	1	0	0	0	16	20,5	0	1	2	1
#56	41	36	1	1	0	0	0	11	22,7	0	1	2	1
#57	40	39	1	1	0	1	0	2	17,5	0	1	2	1
#58	23	46	1	0	1	1	1	5	38,8	0	1	2	1
#59	24	44	0	1	0	0	0	15	31,9	0	1	2	1
#60	40	56	1	1	1	1	0	10	78,4	0	1	2	1
#61	47	24	1	0	0	0	0	2	11,5	0	1	2	1
#62	31	52	1	1	1	1	0	1	13,7	0	1	2	1
#63	45	60	1	0	1	1	0	0	9,2	0	1	3	1
#64	46	39	0	0	0	0	0	10	23,0	0	1	3	1
#65	23	38	1	0	1	1	0	6	14,2	0	1	3	0
#66	30	20	1	0	0	0	0	0	1,6	0	1	3	0

**Tabela B.3 - Cont.**

Receptor	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13
#67	39	53	0	1	1	1	0	5	4,9	0	1	3	0
#68	37	32	1	0	1	0	0	12	34,9	0	1	3	0
#69	32	29	1	0	0	0	0	9	14,5	0	1	3	0
#70	51	46	1	1	1	1	0	6	37,0	0	1	3	0
#71	21	32	1	1	1	1	0	3	5,2	1	1	2	1
#72	7	28	1	1	0	0	0	5	12,4	0	1	2	1
#73	25	47	0	1	0	0	0	3	17,2	0	1	2	1
#74	15	49	0	1	1	1	0	10	7,7	0	1	2	1
#75	32	28	1	1	1	1	0	10	16,8	0	1	2	1
#76	30	31	1	0	0	0	0	0	1,8	0	1	2	1
#77	38	41	0	0	1	1	0	3	21,8	0	1	2	1
#78	28	54	0	0	0	0	0	3	17,1	0	1	2	1
#79	20	52	1	1	0	0	0	4	26,9	0	1	2	1
#80	57	29	1	0	0	1	0	3	7,8	0	1	2	1
#81	27	51	1	1	0	0	0	1	13,9	0	1	2	1
#82	56	53	0	1	1	1	0	5	5,8	0	1	2	1
#83	46	49	1	1	1	1	0	0	8,3	0	1	2	0
#84	45	21	1	0	1	1	0	2	4,5	0	1	2	1
#85	56	27	1	1	1	1	0	2	5,4	0	1	2	1
#86	66	31	1	1	1	1	0	0	16,6	0	1	2	1
#87	49	26	1	1	1	1	0	0	3,7	0	1	2	1
#88	41	52	1	0	0	0	0	4	4,2	0	1	3	1
#89	30	27	1	0	0	0	0	2	5,6	0	1	3	0
#90	24	19	1	1	1	1	0	11	35,4	0	1	3	0
#91	38	32	1	0	0	0	0	0	5,9	0	1	3	0
#92	22	38	1	1	0	0	0	0	17,3	0	1	3	1
#93	23	45	1	1	0	0	0	3	9,9	0	1	3	1
#94	42	41	0	0	1	1	0	20	27,6	0	1	3	0
#95	27	24	1	0	0	0	0	0	10,8	0	1	3	1
#96	36	38	0	1	0	0	0	15	38,2	0	1	3	0
#97	30	42	0	1	0	0	0	1	30,8	0	1	3	0

**Tabela B.3 - Cont.**

Receptor	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13
#98	17	32	1	1	0	0	0	2	0,3	0	1	3	1
#99	38	36	1	1	1	1	0	0	6,1	0	1	3	0
#100	35	32	0	0	1	1	0	11	17,7	0	1	3	0
#101	29	27	1	1	0	0	0	1	16,7	0	1	3	0
#102	32	30	0	1	0	0	0	10	27,8	0	1	3	1
#103	26	33	1	0	1	1	0	2	5,9	0	1	3	0
#104	20	22	1	0	0	0	0	0	11,2	1	1	3	1
#105	52	39	1	1	0	0	0	0	54,4	0	1	3	0
#106	39	33	1	1	0	0	0	0	33,0	0	1	3	0
#107	30	33	1	1	0	0	0	0	14,1	0	1	3	0
#108	36	32	1	1	1	1	0	3	9,6	0	1	3	0
#109	20	48	1	0	0	0	0	4	6,6	0	1	2	1

**Tabela B.4 – Variável dependente *Sobrevida do Enxerto (SE)*, para os 109 receptores, em meses.**

Receptor	SE	Receptor	SE	Receptor	SE	Receptor	SE	Receptor	SE
#1	108,6	#23	97,4	#45	43,9	#67	73,0	#89	1,6
#2	47,7	#24	97,4	#46	73,0	#68	73,0	#90	48,7
#3	4,2	#25	5,6	#47	26,2	#69	73,0	#91	48,7
#4	98,6	#26	51,3	#48	85,2	#70	60,9	#92	48,7
#5	30,1	#27	97,4	#49	19,7	#71	73,0	#93	31,3
#6	0,1	#28	36,8	#50	60,9	#72	60,2	#94	48,7
#7	46,1	#29	102,3	#51	79,2	#73	11,7	#95	48,7
#8	2,4	#30	97,4	#52	85,2	#74	73,0	#96	51,8
#9	98,2	#31	97,4	#53	85,2	#75	0,0	#97	48,7
#10	66,9	#32	97,4	#54	0,1	#76	60,9	#98	48,7
#11	109,6	#33	97,4	#55	85,2	#77	60,9	#99	36,5
#12	8,5	#34	18,4	#56	0,7	#78	78,3	#100	36,1
#13	66,2	#35	61,9	#57	31,6	#79	21,0	#101	0,0
#14	109,6	#36	52,0	#58	13,2	#80	60,9	#102	24,3
#15	109,6	#37	85,2	#59	40,1	#81	5,4	#103	24,3
#16	4,9	#38	97,4	#60	85,2	#82	24,4	#104	12,2

#17	48,2	#39	11,5	#61	46,4	#83	45,4	#105	0,0
#18	109,6	#40	146,1	#62	73,0	#84	23,8	#106	12,2
#19	0,0	#41	97,4	#63	85,2	#85	32,2	#107	12,2
#20	109,6	#42	48,7	#64	85,2	#86	26,0	#108	12,2
#21	0,0	#43	73,0	#65	67,0	#87	0,0	#109	12,2
#22	73,6	#44	5,0	#66	86,7	#88	60,9		

### B.3 – Variáveis padronizadas

**Tabela B.5 – Valores das variáveis predictoras padronizadas segundo equação IV.1.**

Recep	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13
#1	0,53	0,20	0,9	0,10	0,10	0,90	0,10	0,22	0,23	0,10	0,90	0,10	0,90
#2	0,43	0,27	0,9	0,90	0,90	0,90	0,10	0,53	0,30	0,10	0,90	0,10	0,10
#3	0,43	0,18	0,1	0,10	0,90	0,90	0,10	0,41	0,25	0,90	0,90	0,10	0,10
#4	0,54	0,40	0,1	0,10	0,10	0,10	0,10	0,25	0,32	0,10	0,90	0,10	0,10
#5	0,23	0,12	0,9	0,90	0,90	0,90	0,10	0,90	0,27	0,10	0,90	0,50	0,90
#6	0,57	0,37	0,9	0,90	0,10	0,10	0,10	0,19	0,44	0,10	0,90	0,50	0,10
#7	0,49	0,77	0,9	0,10	0,10	0,10	0,10	0,22	0,23	0,10	0,90	0,50	0,10
#8	0,47	0,42	0,9	0,90	0,90	0,90	0,10	0,28	0,15	0,10	0,90	0,50	0,10
#9	0,47	0,90	0,9	0,10	0,10	0,10	0,10	0,35	0,37	0,10	0,90	0,50	0,10
#10	0,14	0,38	0,9	0,90	0,90	0,90	0,10	0,38	0,24	0,10	0,90	0,50	0,90
#11	0,46	0,73	0,9	0,90	0,90	0,90	0,10	0,35	0,23	0,10	0,90	0,50	0,90
#12	0,71	0,70	0,9	0,90	0,90	0,90	0,10	0,35	0,63	0,10	0,90	0,50	0,90
#13	0,15	0,62	0,1	0,90	0,10	0,10	0,10	0,25	0,21	0,10	0,90	0,50	0,90
#14	0,57	0,57	0,9	0,10	0,90	0,90	0,10	0,25	0,20	0,10	0,90	0,50	0,90
#15	0,38	0,20	0,9	0,10	0,90	0,90	0,10	0,28	0,16	0,10	0,90	0,50	0,90
#16	0,73	0,35	0,1	0,10	0,90	0,90	0,10	0,35	0,52	0,10	0,90	0,50	0,90
#17	0,51	0,28	0,1	0,90	0,90	0,90	0,10	0,90	0,40	0,10	0,90	0,50	0,90
#18	0,46	0,30	0,1	0,90	0,10	0,10	0,10	0,41	0,49	0,10	0,90	0,90	0,10
#19	0,78	0,75	0,1	0,90	0,10	0,10	0,10	0,35	0,35	0,10	0,90	0,90	0,10
#20	0,75	0,58	0,1	0,10	0,10	0,10	0,10	0,50	0,26	0,10	0,90	0,90	0,10
#21	0,51	0,32	0,9	0,10	0,90	0,90	0,10	0,22	0,53	0,10	0,90	0,90	0,10
#22	0,66	0,52	0,1	0,10	0,90	0,90	0,10	0,35	0,36	0,10	0,90	0,90	0,10
#23	0,30	0,10	0,9	0,10	0,10	0,10	0,10	0,38	0,56	0,10	0,90	0,50	0,90

**Tabela B.5 – Cont.**

<b>Recep</b>	<b>E1</b>	<b>E2</b>	<b>E3</b>	<b>E4</b>	<b>E5</b>	<b>E6</b>	<b>E7</b>	<b>E8</b>	<b>E9</b>	<b>E10</b>	<b>E11</b>	<b>E12</b>	<b>E13</b>
#24	0,57	0,43	0,9	0,90	0,90	0,90	0,10	0,75	0,30	0,10	0,90	0,50	0,90
#25	0,53	0,48	0,1	0,10	0,10	0,10	0,10	0,16	0,36	0,10	0,90	0,50	0,90
#26	0,33	0,60	0,9	0,10	0,10	0,10	0,10	0,22	0,28	0,10	0,90	0,50	0,90
#27	0,69	0,62	0,1	0,90	0,90	0,90	0,10	0,50	0,22	0,10	0,90	0,50	0,90
#28	0,15	0,55	0,1	0,90	0,10	0,10	0,10	0,32	0,29	0,10	0,90	0,50	0,90
#29	0,42	0,22	0,9	0,10	0,10	0,10	0,10	0,59	0,34	0,10	0,90	0,50	0,90
#30	0,53	0,28	0,9	0,10	0,90	0,90	0,10	0,25	0,28	0,10	0,90	0,50	0,90
#31	0,10	0,25	0,1	0,90	0,10	0,90	0,10	0,72	0,19	0,10	0,90	0,50	0,90
#32	0,47	0,15	0,9	0,90	0,10	0,90	0,10	0,16	0,40	0,10	0,90	0,50	0,90
#33	0,70	0,52	0,9	0,10	0,10	0,10	0,10	0,56	0,61	0,10	0,90	0,50	0,90
#34	0,53	0,45	0,9	0,10	0,10	0,10	0,10	0,16	0,30	0,10	0,90	0,50	0,90
#35	0,17	0,37	0,1	0,90	0,90	0,90	0,10	0,19	0,23	0,10	0,90	0,50	0,90
#36	0,23	0,52	0,1	0,10	0,90	0,90	0,10	0,53	0,14	0,10	0,90	0,50	0,90
#37	0,45	0,60	0,9	0,90	0,10	0,10	0,90	0,84	0,14	0,10	0,90	0,50	0,90
#38	0,21	0,43	0,1	0,90	0,10	0,10	0,10	0,38	0,45	0,10	0,90	0,50	0,90
#39	0,67	0,18	0,9	0,10	0,90	0,90	0,10	0,53	0,23	0,10	0,90	0,50	0,90
#40	0,33	0,58	0,1	0,90	0,90	0,90	0,10	0,16	0,46	0,10	0,90	0,50	0,90
#41	0,14	0,28	0,1	0,90	0,90	0,90	0,10	0,22	0,20	0,10	0,90	0,50	0,90
#42	0,33	0,30	0,1	0,90	0,10	0,10	0,10	0,32	0,36	0,10	0,90	0,50	0,90
#43	0,23	0,57	0,1	0,90	0,10	0,10	0,10	0,41	0,28	0,10	0,90	0,50	0,90
#44	0,19	0,37	0,1	0,90	0,10	0,10	0,10	0,13	0,13	0,10	0,90	0,50	0,90
#45	0,37	0,67	0,9	0,10	0,90	0,90	0,10	0,38	0,24	0,10	0,90	0,50	0,90
#46	0,34	0,23	0,9	0,90	0,10	0,10	0,10	0,56	0,38	0,10	0,90	0,50	0,90
#47	0,69	0,20	0,1	0,90	0,10	0,10	0,10	0,16	0,33	0,10	0,90	0,50	0,90
#48	0,51	0,27	0,1	0,90	0,90	0,90	0,10	0,75	0,37	0,10	0,90	0,50	0,90
#49	0,30	0,62	0,9	0,90	0,10	0,10	0,10	0,25	0,62	0,10	0,90	0,50	0,90
#50	0,63	0,50	0,9	0,10	0,10	0,10	0,10	0,22	0,39	0,10	0,90	0,50	0,90
#51	0,35	0,57	0,9	0,90	0,10	0,10	0,10	0,19	0,20	0,10	0,90	0,50	0,90
#52	0,45	0,30	0,9	0,90	0,90	0,90	0,10	0,28	0,43	0,10	0,90	0,50	0,90
#53	0,54	0,22	0,9	0,90	0,90	0,90	0,10	0,78	0,38	0,10	0,90	0,50	0,90
#54	0,87	0,78	0,9	0,90	0,90	0,90	0,10	0,19	0,29	0,10	0,90	0,50	0,90

**Tabela B.5 – Cont.**

<b>Recep</b>	<b>E1</b>	<b>E2</b>	<b>E3</b>	<b>E4</b>	<b>E5</b>	<b>E6</b>	<b>E7</b>	<b>E8</b>	<b>E9</b>	<b>E10</b>	<b>E11</b>	<b>E12</b>	<b>E13</b>
#55	0,47	0,37	0,9	0,90	0,10	0,10	0,10	0,59	0,31	0,10	0,90	0,50	0,90
#56	0,57	0,38	0,9	0,90	0,10	0,10	0,10	0,44	0,33	0,10	0,90	0,50	0,90
#57	0,55	0,43	0,9	0,90	0,10	0,90	0,10	0,16	0,28	0,10	0,90	0,50	0,90
#58	0,33	0,55	0,9	0,10	0,90	0,90	0,90	0,25	0,50	0,10	0,90	0,50	0,90
#59	0,34	0,52	0,1	0,90	0,10	0,10	0,10	0,56	0,43	0,10	0,90	0,50	0,90
#60	0,55	0,72	0,9	0,90	0,90	0,90	0,10	0,41	0,90	0,10	0,90	0,50	0,90
#61	0,65	0,18	0,9	0,10	0,10	0,10	0,10	0,16	0,22	0,10	0,90	0,50	0,90
#62	0,43	0,65	0,9	0,90	0,90	0,90	0,10	0,13	0,24	0,10	0,90	0,50	0,90
#63	0,62	0,78	0,9	0,10	0,90	0,90	0,10	0,10	0,19	0,10	0,90	0,90	0,90
#64	0,63	0,43	0,1	0,10	0,10	0,10	0,10	0,41	0,34	0,10	0,90	0,90	0,90
#65	0,33	0,42	0,9	0,10	0,90	0,90	0,10	0,28	0,25	0,10	0,90	0,90	0,10
#66	0,42	0,12	0,9	0,10	0,10	0,10	0,10	0,10	0,12	0,10	0,90	0,90	0,10
#67	0,54	0,67	0,1	0,90	0,90	0,90	0,10	0,25	0,15	0,10	0,90	0,90	0,10
#68	0,51	0,32	0,9	0,10	0,90	0,10	0,10	0,47	0,46	0,10	0,90	0,90	0,10
#69	0,45	0,27	0,9	0,10	0,10	0,10	0,10	0,38	0,25	0,10	0,90	0,90	0,10
#70	0,70	0,55	0,9	0,90	0,90	0,90	0,10	0,28	0,48	0,10	0,90	0,90	0,10
#71	0,30	0,32	0,9	0,90	0,90	0,90	0,10	0,19	0,15	0,90	0,90	0,50	0,90
#72	0,11	0,25	0,9	0,90	0,10	0,10	0,10	0,25	0,23	0,10	0,90	0,50	0,90
#73	0,35	0,57	0,1	0,90	0,10	0,10	0,10	0,19	0,28	0,10	0,90	0,50	0,90
#74	0,22	0,60	0,1	0,90	0,90	0,90	0,10	0,41	0,18	0,10	0,90	0,50	0,90
#75	0,45	0,25	0,9	0,90	0,90	0,90	0,10	0,41	0,27	0,10	0,90	0,50	0,90
#76	0,42	0,30	0,9	0,10	0,10	0,10	0,10	0,10	0,12	0,10	0,90	0,50	0,90
#77	0,53	0,47	0,1	0,10	0,90	0,90	0,10	0,19	0,32	0,10	0,90	0,50	0,90
#78	0,39	0,68	0,1	0,10	0,10	0,10	0,10	0,19	0,28	0,10	0,90	0,50	0,90
#79	0,29	0,65	0,9	0,90	0,10	0,10	0,10	0,22	0,38	0,10	0,90	0,50	0,90
#80	0,78	0,27	0,9	0,10	0,10	0,90	0,10	0,19	0,18	0,10	0,90	0,50	0,90
#81	0,38	0,63	0,9	0,90	0,10	0,10	0,10	0,13	0,24	0,10	0,90	0,50	0,90
#82	0,77	0,67	0,1	0,90	0,90	0,90	0,10	0,25	0,16	0,10	0,90	0,50	0,90
#83	0,63	0,60	0,9	0,90	0,90	0,90	0,10	0,10	0,19	0,10	0,90	0,50	0,10
#84	0,62	0,13	0,9	0,10	0,90	0,90	0,10	0,16	0,15	0,10	0,90	0,50	0,90
#85	0,77	0,23	0,9	0,90	0,90	0,90	0,10	0,16	0,16	0,10	0,90	0,50	0,90

**Tabela B.5 – Cont.**

<b>Recep</b>	<b>E1</b>	<b>E2</b>	<b>E3</b>	<b>E4</b>	<b>E5</b>	<b>E6</b>	<b>E7</b>	<b>E8</b>	<b>E9</b>	<b>E10</b>	<b>E11</b>	<b>E12</b>	<b>E13</b>
#86	0,90	0,30	0,9	0,90	0,90	0,90	0,10	0,10	0,27	0,10	0,90	0,50	0,90
#87	0,67	0,22	0,9	0,90	0,90	0,90	0,10	0,10	0,14	0,10	0,90	0,50	0,90
#88	0,57	0,65	0,9	0,10	0,10	0,10	0,10	0,22	0,14	0,10	0,90	0,90	0,90
#89	0,42	0,23	0,9	0,10	0,10	0,10	0,10	0,16	0,16	0,10	0,90	0,90	0,10
#90	0,34	0,10	0,9	0,90	0,90	0,90	0,10	0,44	0,46	0,10	0,90	0,90	0,10
#91	0,53	0,32	0,9	0,10	0,10	0,10	0,10	0,10	0,16	0,10	0,90	0,90	0,10
#92	0,31	0,42	0,9	0,90	0,10	0,10	0,10	0,10	0,28	0,10	0,90	0,90	0,90
#93	0,33	0,53	0,9	0,90	0,10	0,10	0,10	0,19	0,20	0,10	0,90	0,90	0,90
#94	0,58	0,47	0,1	0,10	0,90	0,90	0,10	0,72	0,38	0,10	0,90	0,90	0,10
#95	0,38	0,18	0,9	0,10	0,10	0,10	0,10	0,10	0,21	0,10	0,90	0,90	0,90
#96	0,50	0,42	0,1	0,90	0,10	0,10	0,10	0,56	0,49	0,10	0,90	0,90	0,10
#97	0,42	0,48	0,1	0,90	0,10	0,10	0,10	0,13	0,42	0,10	0,90	0,90	0,10
#98	0,25	0,32	0,9	0,90	0,10	0,10	0,10	0,16	0,10	0,10	0,90	0,90	0,90
#99	0,53	0,38	0,9	0,90	0,90	0,90	0,10	0,10	0,16	0,10	0,90	0,90	0,10
#100	0,49	0,32	0,1	0,10	0,90	0,90	0,10	0,44	0,28	0,10	0,90	0,90	0,10
#101	0,41	0,23	0,9	0,90	0,10	0,10	0,10	0,13	0,27	0,10	0,90	0,90	0,10
#102	0,45	0,28	0,1	0,90	0,10	0,10	0,10	0,41	0,39	0,10	0,90	0,90	0,90
#103	0,37	0,33	0,9	0,10	0,90	0,90	0,10	0,16	0,16	0,10	0,90	0,90	0,10
#104	0,29	0,15	0,9	0,10	0,10	0,10	0,10	0,10	0,21	0,90	0,90	0,90	0,90
#105	0,71	0,43	0,9	0,90	0,10	0,10	0,10	0,10	0,66	0,10	0,90	0,90	0,10
#106	0,54	0,33	0,9	0,90	0,10	0,10	0,10	0,10	0,44	0,10	0,90	0,90	0,10
#107	0,42	0,33	0,9	0,90	0,10	0,10	0,10	0,10	0,24	0,10	0,90	0,90	0,10
#108	0,50	0,32	0,9	0,90	0,90	0,90	0,10	0,19	0,20	0,10	0,90	0,90	0,10
#109	0,29	0,58	0,9	0,10	0,10	0,10	0,10	0,22	0,17	0,10	0,90	0,50	0,90

**Tabela B.6** – Variáveis preditoras após análise estratificada e padronização (as variáveis contínuas não estão saturadas no 95º percentil) segundo equação IV.1. Os receptores que possuem as variáveis *Tipo de Diálise* com valor 1 e *Compatibilidade HLA* com código 0 (doador não relacionado) estão em negrito e negrito itálico, respectivamente. A variável dependente *Sobrevida do Enxerto* é mostrada na Tabela IV.4.

Recep	E1	E2	E3	E4	E5	E6	E8	E9	E10	E12	E13
#1	0,53	0,20	0,9	0,10	0,10	0,90	0,22	0,23	0,10	<b>0,10</b>	0,10
#2	0,43	0,27	0,9	0,90	0,90	0,90	0,53	0,30	0,10	<b>0,10</b>	0,10
#3	0,43	0,18	0,1	0,10	0,90	0,90	0,41	0,25	<b>0,90</b>	<b>0,10</b>	0,10
#4	0,54	0,40	0,1	0,10	0,10	0,10	0,25	0,32	0,10	<b>0,10</b>	0,10
#5	0,23	0,12	0,9	0,90	0,90	0,90	0,90	0,27	0,10	0,50	0,10
#6	0,57	0,37	0,9	0,90	0,10	0,10	0,19	0,44	0,10	0,50	0,10
#7	0,49	0,77	0,9	0,10	0,10	0,10	0,22	0,23	0,10	0,50	0,10
#8	0,47	0,42	0,9	0,90	0,90	0,90	0,28	0,15	0,10	0,50	0,10
#9	0,47	0,90	0,9	0,10	0,10	0,10	0,35	0,37	0,10	0,50	0,10
#10	0,14	0,38	0,9	0,90	0,90	0,90	0,38	0,24	0,10	0,50	0,10
#11	0,46	0,73	0,9	0,90	0,90	0,90	0,35	0,23	0,10	0,50	0,90
#12	0,71	0,70	0,9	0,90	0,90	0,90	0,35	0,63	0,10	0,50	0,90
#13	0,15	0,62	0,1	0,90	0,10	0,10	0,25	0,21	0,10	0,50	0,90
#14	0,57	0,57	0,9	0,10	0,90	0,90	0,25	0,20	0,10	0,50	0,90
#15	0,38	0,20	0,9	0,10	0,90	0,90	0,28	0,16	0,10	0,50	0,90
#16	0,73	0,35	0,1	0,10	0,90	0,90	0,35	0,52	0,10	0,50	0,90
#17	0,51	0,28	0,1	0,90	0,90	0,90	0,90	0,40	0,10	0,50	0,10
#18	0,46	0,30	0,1	0,90	0,10	0,10	0,41	0,49	0,10	0,90	0,10
#19	0,78	0,75	0,1	0,90	0,10	0,10	0,35	0,35	0,10	0,90	0,10
#20	0,75	0,58	0,1	0,10	0,10	0,10	0,50	0,26	0,10	0,90	0,10
#21	0,51	0,32	0,9	0,10	0,90	0,90	0,22	0,53	0,10	0,90	0,10
#22	0,66	0,52	0,1	0,10	0,90	0,90	0,35	0,36	0,10	0,90	0,10
#23	0,30	0,10	0,9	0,10	0,10	0,10	0,38	0,56	0,10	0,50	0,90
#24	0,57	0,43	0,9	0,90	0,90	0,90	0,75	0,30	0,10	0,50	0,90
#25	0,53	0,48	0,1	0,10	0,10	0,10	0,16	0,36	0,10	0,50	0,90
#26	0,33	0,60	0,9	0,10	0,10	0,10	0,22	0,28	0,10	0,50	0,90
#27	0,69	0,62	0,1	0,90	0,90	0,90	0,50	0,22	0,10	0,50	0,90
#28	0,15	0,55	0,1	0,90	0,10	0,10	0,32	0,29	0,10	0,50	0,90
#29	0,42	0,22	0,9	0,10	0,10	0,10	0,59	0,34	0,10	0,50	0,90

**Tabela B.6 – Cont.**

<b>Recep</b>	<b>E1</b>	<b>E2</b>	<b>E3</b>	<b>E4</b>	<b>E5</b>	<b>E6</b>	<b>E8</b>	<b>E9</b>	<b>E10</b>	<b>E12</b>	<b>E13</b>
#30	0,53	0,28	0,9	0,10	0,90	0,90	0,25	0,28	0,10	0,50	0,90
#31	0,10	0,25	0,1	0,90	0,10	0,90	0,72	0,19	0,10	0,50	0,90
#32	0,47	0,15	0,9	0,90	0,10	0,90	0,16	0,40	0,10	0,50	0,90
#33	0,70	0,52	0,9	0,10	0,10	0,10	0,56	0,61	0,10	0,50	0,90
#34	0,53	0,45	0,9	0,10	0,10	0,10	0,16	0,30	0,10	0,50	0,90
#35	0,17	0,37	0,1	0,90	0,90	0,90	0,19	0,23	0,10	0,50	0,90
#36	0,23	0,52	0,1	0,10	0,90	0,90	0,53	0,14	0,10	0,50	0,90
#37	0,45	0,60	0,9	0,90	0,10	0,10	0,84	0,14	0,10	0,50	0,90
#38	0,21	0,43	0,1	0,90	0,10	0,10	0,38	0,45	0,10	0,50	0,90
#39	0,67	0,18	0,9	0,10	0,90	0,90	0,53	0,23	0,10	0,50	0,90
#40	0,33	0,58	0,1	0,90	0,90	0,90	0,16	0,46	0,10	0,50	0,90
#41	0,14	0,28	0,1	0,90	0,90	0,90	0,22	0,20	0,10	0,50	0,90
#42	0,33	0,30	0,1	0,90	0,10	0,10	0,32	0,36	0,10	0,50	0,90
#43	0,23	0,57	0,1	0,90	0,10	0,10	0,41	0,28	0,10	0,50	0,90
#44	0,19	0,37	0,1	0,90	0,10	0,10	0,13	0,13	0,10	0,50	0,90
#45	0,37	0,67	0,9	0,10	0,90	0,90	0,38	0,24	0,10	0,50	0,90
#46	0,34	0,23	0,9	0,90	0,10	0,10	0,56	0,38	0,10	0,50	0,90
#47	0,69	0,20	0,1	0,90	0,10	0,10	0,16	0,33	0,10	0,50	0,90
#48	0,51	0,27	0,1	0,90	0,90	0,90	0,75	0,37	0,10	0,50	0,90
#49	0,30	0,62	0,9	0,90	0,10	0,10	0,25	0,62	0,10	0,50	0,90
#50	0,63	0,50	0,9	0,10	0,10	0,10	0,22	0,39	0,10	0,50	0,90
#51	0,35	0,57	0,9	0,90	0,10	0,10	0,19	0,20	0,10	0,50	0,90
#52	0,45	0,30	0,9	0,90	0,90	0,90	0,28	0,43	0,10	0,50	0,90
#53	0,54	0,22	0,9	0,90	0,90	0,90	0,78	0,38	0,10	0,50	0,90
#54	0,87	0,78	0,9	0,90	0,90	0,90	0,19	0,29	0,10	0,50	0,90
#55	0,47	0,37	0,9	0,90	0,10	0,10	0,59	0,31	0,10	0,50	0,90
#56	0,57	0,38	0,9	0,90	0,10	0,10	0,44	0,33	0,10	0,50	0,90
#57	0,55	0,43	0,9	0,90	0,10	0,90	0,16	0,28	0,10	0,50	0,90
#58	0,33	0,55	0,9	0,10	0,90	0,90	0,25	0,50	0,10	0,50	0,90
#59	0,34	0,52	0,1	0,90	0,10	0,10	0,56	0,43	0,10	0,50	0,90
#60	0,55	0,72	0,9	0,90	0,90	0,90	0,41	0,90	0,10	0,50	0,90
#61	0,65	0,18	0,9	0,10	0,10	0,10	0,16	0,22	0,10	0,50	0,90

**Tabela B.6 – Cont.**

<b>Recep</b>	<b>E1</b>	<b>E2</b>	<b>E3</b>	<b>E4</b>	<b>E5</b>	<b>E6</b>	<b>E8</b>	<b>E9</b>	<b>E10</b>	<b>E12</b>	<b>E13</b>
#62	0,43	0,65	0,9	0,90	0,90	0,90	0,13	0,24	0,10	0,50	0,90
#63	0,62	0,78	0,9	0,10	0,90	0,90	0,10	0,19	0,10	0,90	0,90
#64	0,63	0,43	0,1	0,10	0,10	0,10	0,41	0,34	0,10	0,90	0,90
#65	0,33	0,42	0,9	0,10	0,90	0,90	0,28	0,25	0,10	0,90	0,90
#66	0,42	0,12	0,9	0,10	0,10	0,10	0,10	0,12	0,10	0,90	0,90
#67	0,54	0,67	0,1	0,90	0,90	0,90	0,25	0,15	0,10	0,90	0,90
#68	0,51	0,32	0,9	0,10	0,90	0,10	0,47	0,46	0,10	0,90	0,10
#69	0,45	0,27	0,9	0,10	0,10	0,10	0,38	0,25	0,10	0,90	0,10
#70	0,70	0,55	0,9	0,90	0,90	0,90	0,28	0,48	0,10	0,90	0,10
#71	0,30	0,32	0,9	0,90	0,90	0,90	0,19	0,15	<b>0,90</b>	0,50	0,90
#72	0,11	0,25	0,9	0,90	0,10	0,10	0,25	0,23	0,10	0,50	0,90
#73	0,35	0,57	0,1	0,90	0,10	0,10	0,19	0,28	0,10	0,50	0,90
#74	0,22	0,60	0,1	0,90	0,90	0,90	0,41	0,18	0,10	0,50	0,90
#75	0,45	0,25	0,9	0,90	0,90	0,90	0,41	0,27	0,10	0,50	0,90
#76	0,42	0,30	0,9	0,10	0,10	0,10	0,10	0,12	0,10	0,50	0,90
#77	0,53	0,47	0,1	0,10	0,90	0,90	0,19	0,32	0,10	0,50	0,90
#78	0,39	0,68	0,1	0,10	0,10	0,10	0,19	0,28	0,10	0,50	0,90
#79	0,29	0,65	0,9	0,90	0,10	0,10	0,22	0,38	0,10	0,50	0,90
#80	0,78	0,27	0,9	0,10	0,10	0,90	0,19	0,18	0,10	0,50	0,90
#81	0,38	0,63	0,9	0,90	0,10	0,10	0,13	0,24	0,10	0,50	0,90
#82	0,77	0,67	0,1	0,90	0,90	0,90	0,25	0,16	0,10	0,50	0,90
#83	0,63	0,60	0,9	0,90	0,90	0,90	0,10	0,19	0,10	0,50	0,10
#84	0,62	0,13	0,9	0,10	0,90	0,90	0,16	0,15	0,10	0,50	0,90
#85	0,77	0,23	0,9	0,90	0,90	0,90	0,16	0,16	0,10	0,50	0,90
#86	0,90	0,30	0,9	0,90	0,90	0,90	0,10	0,27	0,10	0,50	0,90
#87	0,67	0,22	0,9	0,90	0,90	0,90	0,10	0,14	0,10	0,50	0,90
#88	0,57	0,65	0,9	0,10	0,10	0,10	0,22	0,14	0,10	0,90	0,90
#89	0,42	0,23	0,9	0,10	0,10	0,10	0,16	0,16	0,10	0,90	0,10
#90	0,34	0,10	0,9	0,90	0,90	0,90	0,44	0,46	0,10	0,90	0,10
#91	0,53	0,32	0,9	0,10	0,10	0,10	0,10	0,16	0,10	0,90	0,10
#92	0,31	0,42	0,9	0,90	0,10	0,10	0,10	0,28	0,10	0,90	0,90
#93	0,33	0,53	0,9	0,90	0,10	0,10	0,19	0,20	0,10	0,90	0,90

**Tabela B.6 – Cont.**

Recep	E1	E2	E3	E4	E5	E6	E8	E9	E10	E12	E13
#94	0,58	0,47	0,1	0,10	0,90	0,90	0,72	0,38	0,10	0,90	0,10
#95	0,38	0,18	0,9	0,10	0,10	0,10	0,10	0,21	0,10	0,90	0,90
#96	0,50	0,42	0,1	0,90	0,10	0,10	0,56	0,49	0,10	0,90	0,10
#97	0,42	0,48	0,1	0,90	0,10	0,10	0,13	0,42	0,10	0,90	0,10
#98	0,25	0,32	0,9	0,90	0,10	0,10	0,16	0,10	0,10	0,90	0,90
#99	0,53	0,38	0,9	0,90	0,90	0,90	0,10	0,16	0,10	0,90	0,10
#100	0,49	0,32	0,1	0,10	0,90	0,90	0,44	0,28	0,10	0,90	0,10
#101	0,41	0,23	0,9	0,90	0,10	0,10	0,13	0,27	0,10	0,90	0,10
#102	0,45	0,28	0,1	0,90	0,10	0,10	0,41	0,39	0,10	0,90	0,90
#103	0,37	0,33	0,9	0,10	0,90	0,90	0,16	0,16	0,10	0,90	0,10
<b>#104</b>	0,29	0,15	0,9	0,10	0,10	0,10	0,10	0,21	<b>0,90</b>	0,90	0,90
#105	0,71	0,43	0,9	0,90	0,10	0,10	0,10	0,66	0,10	0,90	0,10
#106	0,54	0,33	0,9	0,90	0,10	0,10	0,10	0,44	0,10	0,90	0,10
#107	0,42	0,33	0,9	0,90	0,10	0,10	0,10	0,24	0,10	0,90	0,10
#108	0,50	0,32	0,9	0,90	0,90	0,90	0,19	0,20	0,10	0,90	0,10
#109	0,29	0,58	0,9	0,10	0,10	0,10	0,22	0,17	0,10	0,50	0,90

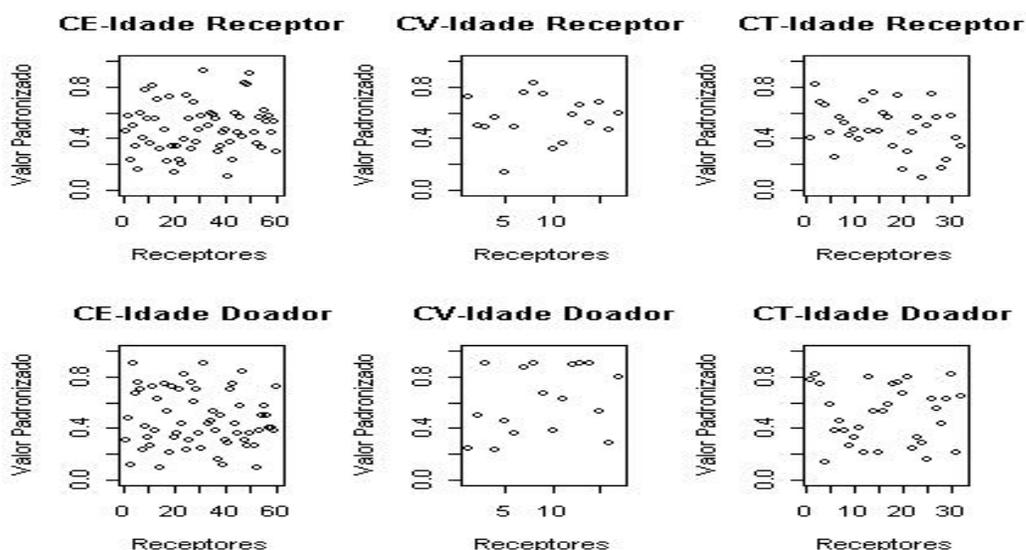
#### B.4 – Conjuntos de Estimação, Validação e Teste

Os diagramas de espalhamento das variáveis preditoras e dependente estão mostrados nas Figuras B.1 a B.6. As variáveis estão padronizadas e saturadas no 95º percentil, sendo que a variável *Sobrevida do Enxerto* está apresentada nas nove faixas, conforme Tabela IV.4.

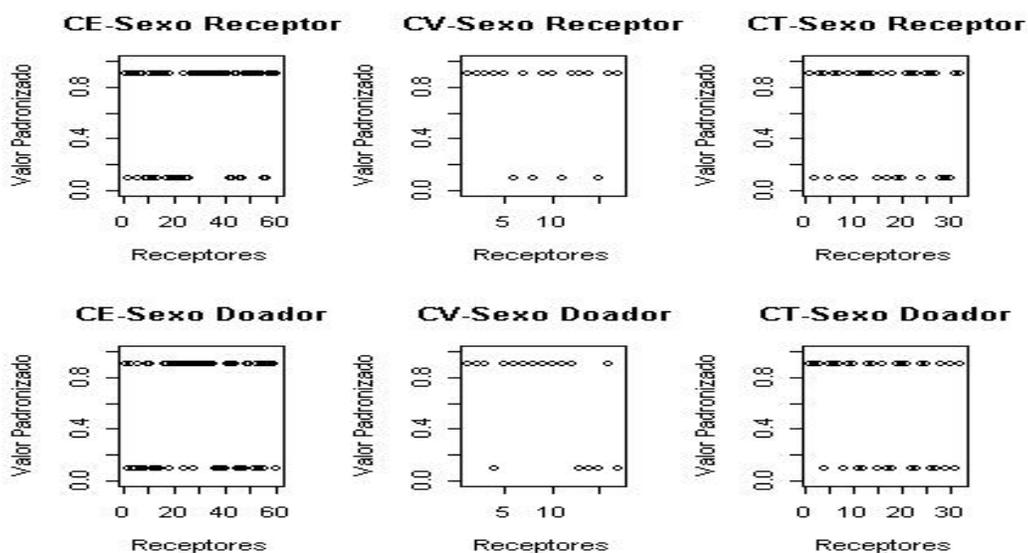
As variáveis contínuas nos três conjuntos são estatisticamente similares em relação à tendência central dos dados pelo teste de Kruskal-Wallis: *Idade do Receptor* (KW=2,97, p=0,23); *Idade do Doador* (KW=4,77, p=0,09); *Transfusões Sangüíneas* (KW=4,22, p=0,12); *Duração da Diálise* (KW=2,02, p=0,36), e *Sobrevida do Enxerto* (KW=0,46, p=0,79). Para homogeneidade das variâncias, também não houve diferenças, exceto para *Transfusões Sangüíneas* (FK=6,4, p=0,04). Dados mostrados no Apêndice D.

Em relação às variáveis dicotômicas e categóricas, as diferenças foram avaliadas verificando-se o teste de comparação de proporções qui-quadrado de Pearson, em tabelas de contingência. Neste caso, a hipótese nula foi de não haver diferença entre

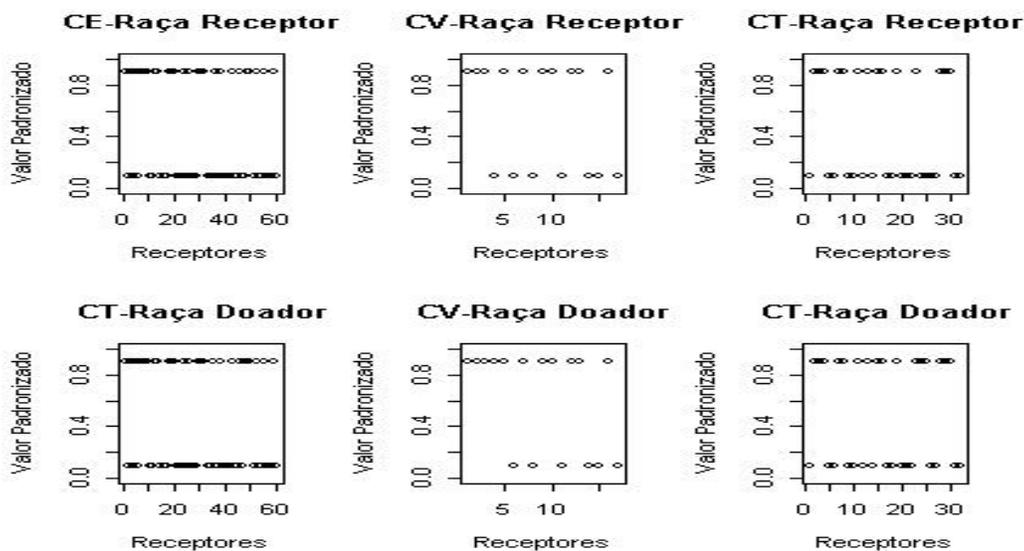
as proporções em cada categoria dentro de CE, CV e CT. Todos os grupos mostraram-se sem heterogeneidades a nível de significância de 5%, com  $p=0,61$ ,  $p=0,59$ ,  $p=0,51$ ,  $p=0,36$ ,  $p=0,63$ ,  $p=0,93$  e  $p=0,94$  para E3, E4, E5, E6, E10, E13 e E12, respectivamente (Tabela B.5 e Apêndice D).



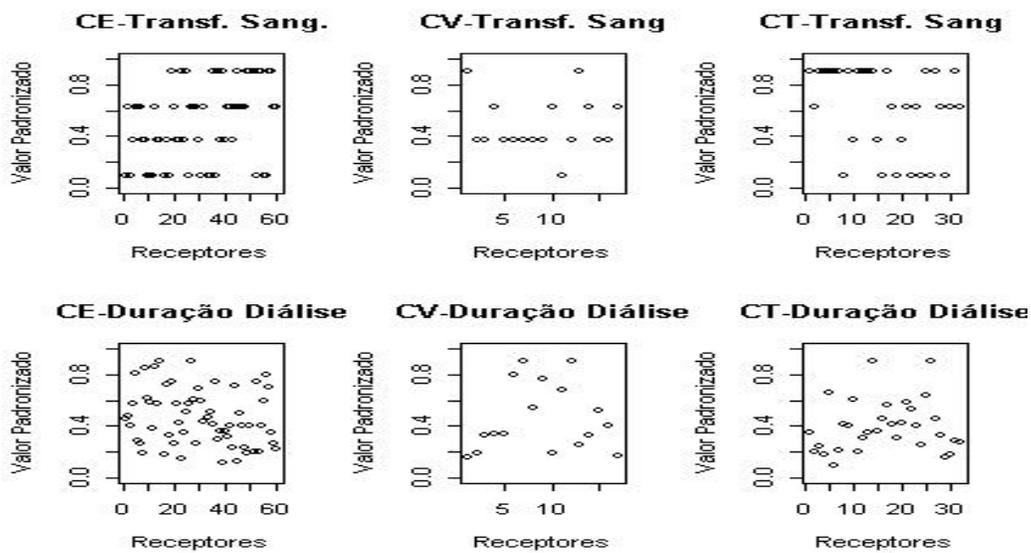
**Figura B.1** – Distribuição dos receptores por conjuntos, para *Idade do Receptor* (linha superior) e *Idade do Doador* (linha inferior).



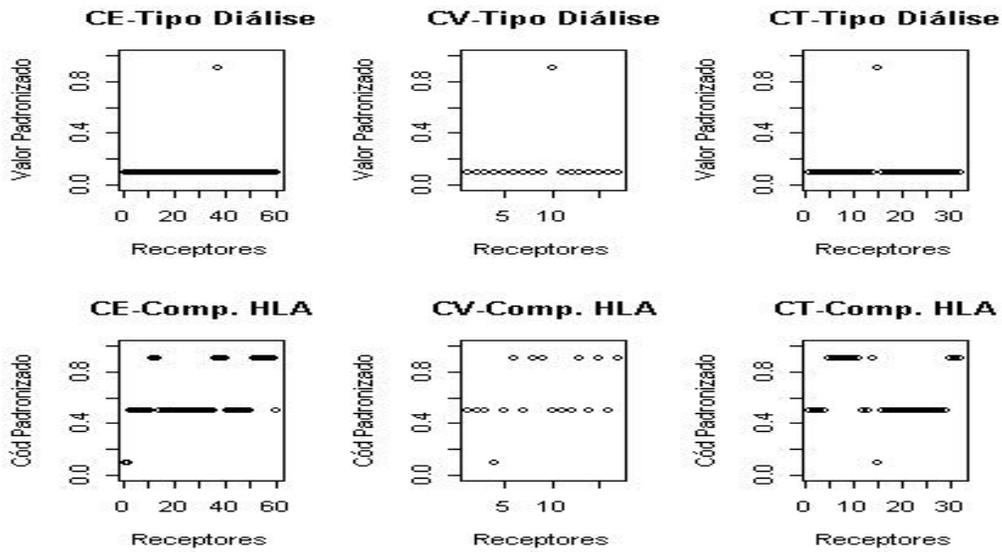
**Figura B.2** – Distribuição dos receptores por Conjunto, para *Sexo do Receptor* (linha superior) e *Sexo do Doador* (linha inferior).



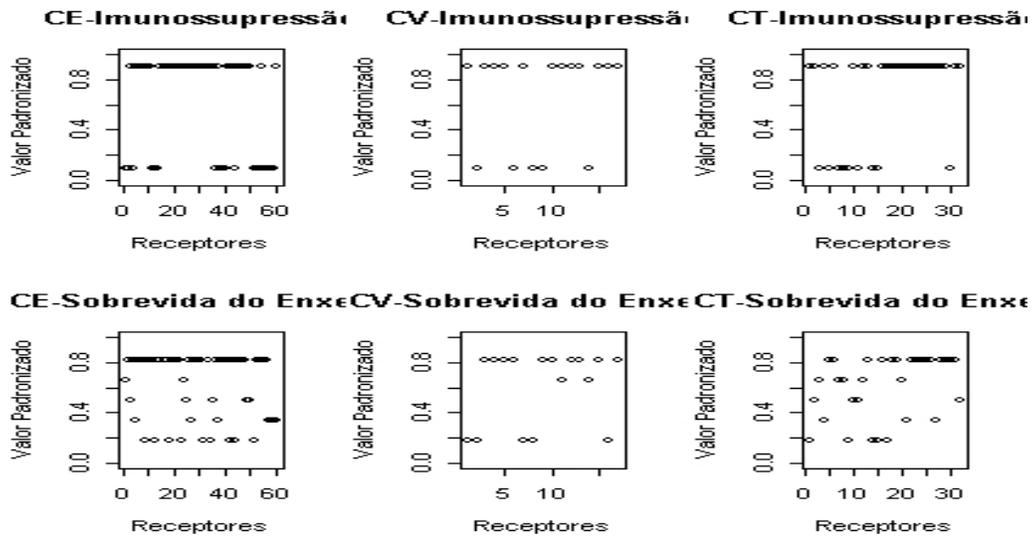
**Figura B.3** – Distribuição dos receptores por Conjuntos, para *Raça do Receptor* (linha superior) e *Raça do Doador* (linha inferior).



**Figura B.4** – Distribuição dos receptores por Conjunto, para *Transfusões Sangüíneas* (linha superior) e *Duração da Diálise* (linha inferior).



**Figura B.5** – Distribuição dos receptores por Conjunto, para *Tipo de Diálise* (linha superior) e *Compatibilidade HLA* (linha inferior).



**Figura B.6** – Distribuição dos receptores por Conjunto, para *Imunossupressão* (linha superior) e *Sobrevida do Enxerto* (linha inferior).

# APÊNDICE C

## BOOTSTRAP

### C.1 – Introdução

Como visto no Capítulo II, a metodologia *Bootstrap* (BST) não-paramétrica possibilita o aumento do poder estatístico do teste quando não se conhece a fdp da população. É um método computacionalmente intensivo, utilizado para aumentar a acurácia de uma determinada estatística de interesse. Utilizando BST no Conjunto de Treinamento, pode ser possível aumentar a capacidade de generalização de uma RNA (ZHANG, 1999, HAHN, 2005).

A utilização de *Bootstrap* pode ocasionar *bias* entre as estatísticas calculadas para a amostra original e a média das estatísticas das amostras *Bootstrap*. Como já dito, a utilização de *Bootstrap* balanceado (BSTbal) visa diminuir o *bias*, uma vez que todos os casos estão presentes na mesma proporção. Como o programa de geração das amostras *Bootstrap* foi feito em LabVIEW®, os resultados obtidos pelo mesmo foram comparados com exemplos extraídos de MANLY (1997), para validação do programa e comparação entre o *Bootstrap* tradicional e o *Bootstrap* balanceado.

### C.2 – Validação do LabVIEW®

Utilizou-se um exemplo extraído de MANLY, 1997 (exemplo 3.1), para a validação na obtenção das estatísticas BST. Seja o vetor

$$X = (3,56;0,69;0,10;1,84;3,93;1,25;0,18;1,13;0,27;0,50;0,67;0,01;0,61; \\ 0,82;1,70;0,39;0,11;1,20;1,21;0,72),$$

de 20 elementos, cujo desvio-padrão ( $\hat{\sigma} = 1,03$ ) é a estatística de interesse. Logo, pretende-se verificar a acurácia do desvio-padrão da amostra em estimar o desvio-padrão da população, ou seja, determinar o *bias* da estatística escolhida, definida como a diferença entre a estatística dada pela simulação BST e a estatística da “população” BST (que, neste caso, é a amostra original). A estatística escolhida das amostras BST é calculada como se segue:

$$\hat{\sigma}_B = B^{-1} \cdot \sum_{i=1}^B \hat{\sigma}_i \quad (C.1)$$

onde  $\hat{\sigma}_B$  é a estatística BST calculada para as  $B$  replicações e  $\hat{\sigma}_i$  é a estatística para cada amostra BST. O desvio-padrão em relação a  $\hat{\sigma}_B$  de cada estatística BST é:

$$DP = \sqrt{(B-1)^{-1} \cdot \sum_{i=1}^B (\hat{\sigma}_i - \hat{\sigma}_B)^2} \quad (C.2)$$

onde DP é o desvio-padrão e os demais termos já definidos na equação (C.1).

A comparação entre os resultados de MANLY (1997) com os BST (para BST tradicional e balanceado) aqui obtidos, ambos com 1000 replicações, é mostrada na Tabela C.1:

**Tabela C.1** - Comparação entre os resultados das simulações para o exemplo apresentado por MANLY (1997) e as simulações utilizando BST tradicional e BST balanceado (BSTbal) em LabVIEW®.

Estatística de Interesse ou Estimador	MAINLY (1997)	LabVIEW® BST	LabVIEW® BSTbal
Média <i>Bootstrap</i> de $\hat{\sigma}$	0,97	0,95	1,00
DP <i>Bootstrap</i> de $\hat{\sigma}$	0,25	0,26	0,26
<i>Bias</i> de $\hat{\sigma}$	-0,06	-0,08	-0,03

Tem-se então que  $\hat{\sigma}_B$  é a média aritmética dos desvios-padrões das amostras BST, DP *Bootstrap* de  $\hat{\sigma}$  é o desvio-padrão de  $\hat{\sigma}_B$ , e o *bias* de  $\hat{\sigma}$  é a diferença  $\hat{\sigma}_B - \hat{\sigma}$ .

Uma parcela das 1000 simulações desenvolvidas em LabVIEW® para o BST tradicional é mostrada na Figura C.1, que pode ser comparada com a Figura C.2, retirada de MANLY, 1997.



O método *Bootstrap* balanceado garante que os elementos amostrais de interesse estarão presentes nas amostras simuladas na mesma proporção. Neste trabalho, é utilizado o *balanceamento de primeira ordem*, e a medida de sua eficiência sobre o método tradicional é a razão das variâncias:

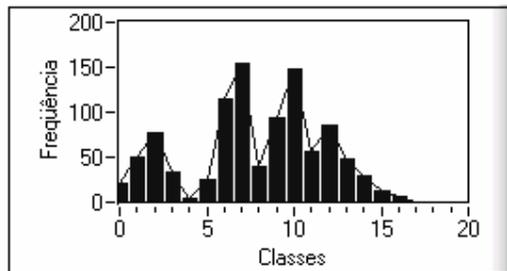
$$E_f = \frac{\text{var}_{\text{trad}}(\hat{\sigma}_B)}{\text{var}_{\text{bal}}(\hat{\sigma}_B)} \quad (\text{C.3})$$

onde os subscritos indicam o método utilizado (DAVINSON e HINKLEY, 1997). A correção do estimador da amostra original pelo bias de BSTbal (Tabela C.1) resulta no estimador do desvio-padrão não tendencioso da amostra original ( $\hat{\sigma} = 1,03 - (-0,03) = 1,06$ ). O método BSTbal, como dito, implica na diminuição do *bias* (DAVINSON e HINKLEY, 1997).

Como na abordagem estatística tradicional é comum a suposição de que a amostra é oriunda de uma população cuja fdp é Gaussiana (ou outra qualquer conhecida), na utilização de *Bootstrap* não-paramétrico tal não acontece, uma vez que é pressuposto que a amostra é demasiadamente pequena para qualquer inferência acerca da população. A abordagem não-paramétrica do BST tem vantagens sobre a paramétrica por não existir a pressuposição de características estatísticas da população, podendo-se ajustar um maior número de modelos e evitando *bias* que poderiam ser introduzidas pela parametrização. Por esta razão, ela foi a utilizada neste trabalho, e, a menos de explicitado em contrário, a menção a *Bootstrap* deve ser entendida como sendo *Bootstrap* não-paramétrico.

Para a determinação do Intervalo de Confiança da estatística *Bootstrap*, é possível utilizar o método de Percentis de Efron (MANLY, 1997). DAVINSON e HINKLEY (1997) também o intitulam de Método do Percentil Básico (MPB). O Método do Percentil Básico pressupõe que haja uma função  $f(\cdot)$  monotonicamente crescente que transforme a fdp da estatística das amostras BST em uma fdp Gaussiana (sem que seja necessário o seu conhecimento). Se  $f(\hat{\theta}_B)$  tem fdp Gaussiana, então o Intervalo de Confiança para  $(1-\alpha).100\%$  será determinado pelos valores externos adjacentes aos percentis de  $\alpha/2$  (limite inferior) e de  $(1-\alpha/2)$  (limite superior), pois os valores ordenados das replicações BST tem a mesma ordem dos valores transformados  $f(\hat{\theta}_B)$ . O Intervalo de Confiança é o intervalo que possui uma probabilidade  $(1-\alpha).100\%$  de conter a estatística de interesse.

Como exemplo, pode-se utilizar o mesmo vetor  $X$ , dado acima, para 1000 replicações BST e verificar a validade do programa LabVIEW® em estimar Intervalos de Confiança. O Histograma de  $\hat{\sigma}_B$  é mostrado na Figura C.3:



**Figura C.3** - Histograma dos valores da estatística, valor mínimo de 0,35 da classe 0, e máximo da classe 19 de 1,63, 1000 amostras BST.

Pode-se ver que a distribuição dos valores das estatísticas BST nem mesmo é simétrica, mas sob o pressuposto de que exista uma transformação tal que a converta em Gaussiana, os limites para o Intervalo de Confiança de 95% serão 0,47 e 1,39 (contra 0,44 e 1,40 encontrados por MANLY). Os limites foram encontrados colocando-se os valores das estatísticas encontradas em cada amostra BST em ordem crescente e observando-se os valores referentes às 24ª e 976ª posições. O programa desenvolvido em LabVIEW obteve resultados similares aos mostrados em MANLY (1997).

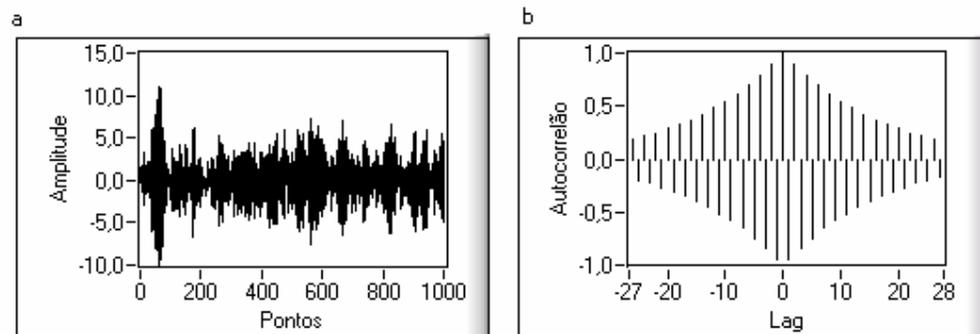
Subseqüentemente, Efron desenvolveu dois métodos para melhorar a estimativa dos Intervalos de Confiança baseados em Percentis de BST, os métodos BC (*Bias Corrected*) e ABC (*Accelerated Bias Corrected*). Outro método também utilizado é o *student-BST* (DAVINSON e HINKLEY, 1997; MANLY, 1997).

### C.3 – Algumas Aplicações do BST

Uma aplicação dos conceitos vistos até aqui foi a descrita por CHANG *et al.* (2002). Os autores utilizaram dois índices de atividade de Lupus Eritomatoso Sistêmico (LES). O *Systemic Lupus Activity Measure revised* (SLAM-R) e o *Systemic Lupus Eritematosus Disease Activity Index* (SLEDAI) foram obtidos a partir de uma

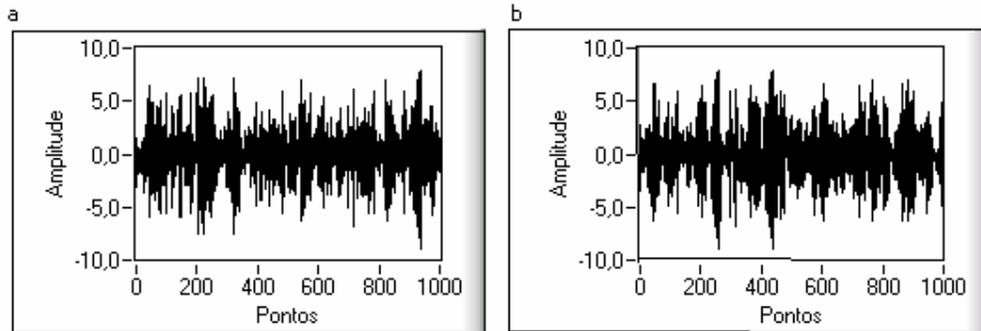
amostra populacional de 76 pacientes, para os quais com informações (dados) são consideradas completas. Como os índices de atividade possuíam fdp desconhecida, foi utilizada simulação BST com o número de replicações igual a 1000 e com uma abordagem de diminuição da variabilidade semelhante à do método BSTbal. Os Intervalos de Confiança foram determinados pelo Método Percentil Básico.

Outra possível utilização dos métodos de Perm e BST é na análise de séries temporais. DAVINSON E HINKLEY (1997) discorrem sobre métodos de Perm e BST utilizando blocos da série original, de tal maneira que  $n = bl$ , onde  $b$  é o número de blocos sem superposição,  $l$  é o comprimento de cada bloco e  $n$  é o número total de pontos da série. Assim, a estrutura da série é mantida nas regiões centrais dos blocos, alterando-se somente no início ou fim dos mesmos, porém não é mais garantida a estacionariedade da série, nem a dependência temporal. A partir da simulação, pode-se fazer inferências sobre a estrutura da série e  $H_0$ ,  $H_1$  e valor  $p$ . Encontra-se na Figura C.4 o exemplo de uma série com estrutura Autoregressiva de ordem 1 (AR(1)), e sua Função de Autocorrelação Normalizada (NFAC):



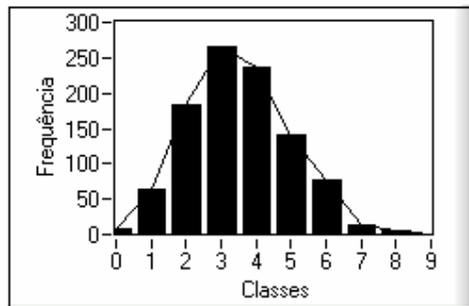
**Figura C.4** – **a)** sinal AR(1),  $y(n)=-0,95.y(n-1)+e(n)$ , onde  $e(n)$  é ruído branco Gaussiano, 1000 pontos; **b)**: NFAC para lag's [-27,28].

A série pode ser dividida em 50 blocos de 20 pontos sem superposição e testar a  $H_0$  de que os pontos da série são iid. Uma estatística que pode ser utilizada é o valor da NFAC para o lag(1), que neste caso é -0,95. Este valor indica uma grande dependência temporal entre amostras consecutivas. Na Figura C.5 é mostrada a primeira e a última série BST geradas, a partir de 999 replicações:



**Figura C.5 – a)** Série da 1ª replicação; **b)** Série da 999ª replicação. O início e o comprimento dos blocos foram obtidos a partir de um gerador de números pseudo-aleatórios com fdp uniforme.

O histograma dos valores do 1º Lag é mostrado na Figura C.6. Não foi determinado o valor médio da estatística porque a intenção é mostrar o grau de confiança em rejeitar  $H_0$ .



**Figura C.6** - Histograma dos valores de Lag(1), para 999 replicações BST. A classe 0 tem valor mínimo de  $-0,94$ , médio de  $-0,93$ , e a classe 9 tem valor máximo de  $-0,84$ , médio de  $-0,85$ .

Pode-se ver que o histograma mostra uma assimetria à esquerda e, mesmo assim, a estatística para a amostra original situa-se fora dos limites das classes 0 e 9, mostrando fortes evidências a favor da Hipótese Alternativa, de que há dependência temporal, com um grau de confiança nesta afirmativa avaliado por:

$$p = \frac{1 + \text{no}\{\hat{\theta}_B \geq \hat{\theta}\}}{B + 1} = \frac{1 + 0}{999 + 1} = 0,001 \quad (\text{C.4})$$

onde  $\text{no}\{\hat{\theta}_B \geq \hat{\theta}\}$  é o número de vezes que a estatística determinada das amostras *Bootstrap* excede o valor da estatística calculada da amostra original, B é o número de replicações BST e  $p$  é o valor P (DAVINSON E HINKLEY, 1997, pág. 158).

A média da estatística nas amostras BST nos dá o valor de  $\text{La}\hat{g}(1)_B = -0,91$ . Podemos ver que este valor ainda indica grande dependência temporal, o que não aconteceria para quantidade de blocos menores. Por exemplo, com  $b = 200$ , o valor da estatística cai para  $-0,76$  ( $-0,81$  a  $-0,71$ ) e não teríamos evidências para rejeitar  $H_0$ .

É evidente que, aumentando-se o tamanho dos blocos, aumenta-se a chance de manutenção da estrutura da série inicial, portanto o método exemplificado acima incorre na dependência do comprimento dos blocos. DAVINSON E HINKLEY (1997) discorrem sobre vários métodos para a determinação do comprimento dos blocos, bem como para manter a estrutura da série original. O método do BST Estacionário (POLITIS e ROMANO, 1994; DAVINSON e HINKLEY, 1997) mantém a estacionariedade da série (quando for este o caso), gerando “pseudo” séries temporais a partir da série original utilizando um modelo paramétrico de fdp geométrica (embora outras distribuições possam ser utilizadas, ver POLITIS e ROMANO, 1994).