



**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA E
ESTATÍSTICA**

Charles Costa de Oliveira

**Aplicação do Modelo Linear Generalizado Binomial e Redes
Neurais Artificiais na Estimativa de Mortalidade de Espécies
Arbóreas.**

Orientadora: Profa. Dra. Terezinha Ferreira de Oliveira

Co-orientadora: Profa. Dra. Marinalva Cardoso Maciel

Belém, 09 de Junho de 2017.

Charles Costa de Oliveira

**Aplicação do Modelo Linear Generalizado Binomial e Redes
Neurais Artificiais na Estimativa de Mortalidade de Espécies
Arbóreas.**

Dissertação de Mestrado apresentada ao programa de Pós-Graduação em Matemática e Estatística da Universidade Federal do Pará (PPGME-UFGPA) como requisito para obtenção do título de Mestre em Estatística. Orientadora: Profa. Dra. Terezinha Ferreira de Oliveira. Co-orientadora: Profa. Dra. Marinalva Cardoso Maciel.

Área de Concentração: Modelos Lineares Generalizados

Orientadora: Profa. Dra. Terezinha Ferreira de Oliveira

Co-orientadora: Profa. Dra. Marinalva Cardoso Maciel

Belém, 09 de Junho de 2017.

Dados Internacionais de Catalogação - na - Publicação (CIP)
Biblioteca de Pós-Graduação do ICEN/UFPA

Oliveira, Charles Costa de

Aplicação do modelo linear generalizado binomial e redes neurais artificiais na estimativa de mortalidade de espécies Arbóreas/ Charles Costa de Oliveira; orientador, Terezinha Ferreira de Oliveira.-2017.

66 f. il. 29 cm

Inclui bibliografias

Dissertação (Mestrado) – Universidade Federal do Pará, Instituto de Ciências Exatas e Naturais, Programa de Pós-Graduação em Matemática e Estatística, Belém, 2017.

1. Teoria da estimativa. 2. Árvores-Mortalidade-Indicadores.
3. Redes neurais. 4. Análise de regressão logística. I. Oliveira, Terezinha Ferreira, orient. II. Título.

CDD – 22 ed. 519.544

Charles Costa de Oliveira

**Aplicação do Modelo Linear Generalizado Binomial e Redes
Neurais Artificiais na Estimativa de Mortalidade de Espécies
Arbóreas.**

Dissertação de Mestrado julgada e aprovada ao programa de Pós-Graduação em Matemática e Estatística da Universidade Federal do Pará (PPGME-UFGPA) como requisito para obtenção do título de Mestre em Estatística.

Data da defesa: 09 de Junho de 2017.

Resultado: APROVADO

João Marcelo Brazão Protázio

Prof. Dr. João Marcelo Brazão Protázio.
(Coordenador do Programa de Pós-Graduação em Matemática e Estatística - PPGME)

Banca Examinadora

Terezinha Ferreira de Oliveira

Profa. Dra Terezinha Ferreira de Oliveira.

Universidade Federal do Pará (PPGME-UFGPA)

Orientadora

Marinalva Cardoso Maciel

Profa. Dra. Marinalva Cardoso Maciel.

Universidade Federal do Pará (FAEST-UFGPA)

Co-Orientadora

João Marcelo Brazão Protázio

Prof. Dr. João Marcelo Brazão Protázio.

Universidade Federal do Pará (PPGME-UFGPA)

Examinador

Agradecimentos

- * A Deus pelo dom da vida e que pela sua graça e misericórdia permitiu que eu passasse no vestibular e agora concluísse este curso superior;
- * À minha avó Maria Silva de Oliveira por sempre olhar e interceder por mim e minha família;
- * À minha mãe Maria Suely por sua dedicação incondicional e fé no meu futuro através dos estudos;
- * À minha namorada Denize Rafaela por está sempre comigo nos bons e maus momentos, por sempre me incentivar a crescer como estatístico;
- * Aos colegas do curso Miguel Trocolis, Tarcilio Roger, Fádua Antonio e Shirlaine Mores;
- * Ao amigo Carlos Takeshi Kudo Yasojima por seu auxílio na obtenção de alguns resultados;
- * À Profa. Terezinha Ferreira de Oliveira, por sua paciência e orientação;
- * À Profa. Marinalva Cardoso Maciel, por sua co-orientação;
- * À EMBRAPA AMAZÔNIA ORIENTAL na pessoa do Doutor Lucas Jose Mazzei de Freitas pela disponibilização das informações que contribuíram para o êxito deste trabalho;
- * Ao Centro de Registro e Indicadores Acadêmicos, na pessoa da diretora Julieta de Andrade Jathahy por sua compreensão e incentivo para a conclusão deste trabalho.
- * A todos que de alguma forma contribuíram para a conclusão deste trabalho.

À minha família, em especial à minha mãe.

*“Deus é o meu rochedo, nele confiarei; o meu escudo,
e a força da minha salvação, o meu alto retiro,
e o meu refúgio”.*

II Samuel 22, 3.

*Os humilhados serão exaltados,
minha vitória tem sabor de mel.*

RESUMO

OLIVEIRA, Charles Costa. Aplicação do Modelo Linear Generalizado Binomial e Redes Neurais Artificiais na Estimativa de Mortalidade de Espécies Arbóreas. 2017. Dissertação de Mestrado. Programa de Pós-Graduação em Matemática e Estatística – PPGME, UFPA, Belém, Pará, Brasil.

O presente estudo teve como objetivo identificar os fatores relacionados à mortalidade de seis espécies arbóreas por meio de um modelo estatístico que associe os indicadores de competição e diversidade a essa mortalidade. As ferramentas utilizadas neste trabalho foram: Modelos Lineares Generalizados em Regressão Logística e Redes Neurais Artificiais, a utilização das técnicas permitiu alcançar os objetivos propostos neste trabalho. Na Regressão Logística foi evidenciado que as variáveis: diâmetro altura do peito, número de unidades amostrais na subparcela, tratamento, espécies e a interação entre o número de unidades amostrais na subparcela e a diversidade são importantes para o estudo da mortalidade das espécies arbóreas, pois seus níveis descritivos de probabilidade foram significativos ao nível de 5%. O desempenho obtido pela Rede Neural Artificial foi de grande relevância, pois se obteve uma melhora na classificação dos dados, sendo um instrumento de grande apoio ao diagnóstico florestal.

Palavras Chaves: Regressão Logística, Redes Neurais Artificiais, Indicadores e Mortalidade de espécies arbóreas.

ABSTRACT

OLIVEIRA, Charles Costa. Application of the Binomial Generalized Linear Model and Artificial Neural Networks in the Estimation of Tree Species Mortality. 2017. Master Thesis. Master's Program in Mathematics and Statistics – PPGME, UFPA, Belém, Pará, Brazil.

The present study had as objective identify factors related to mortality of six tree species through a statistical model which associates competition and diversity indicators to that mortality. The tools used in this search were: Generalized Linear Models in Logistic Regression and Artificial Neural Networks. The use of this techniques allowed to reach the goals proposed in this thesis. In the Logistic Regression has been shown that the variables: diameter at breast height, number of sampling units in subplot, treatment, species and interaction between number of sampling units in subplot and diversity, are important to study of tree species mortality, because its descriptive levels of probability were significant at the level of 5%. The performance achieved for Artificial Neural Networks had great relevance, because it was obtained a significant improvement in the data classification, being a great support tool to forest diagnosis.

Key-words: Logistic Regression, Artificial Neural Networks, Indicators and Mortality Of Tree Species.

ÍNDICE

Resumo.....	vii
Abstract	viii
Lista de tabelas	xii
Lista de figuras	xiii
Lista de abreviaturas e siglas	xiv
1. Introdução.....	1
1.1 Aspectos gerais	1
1.2 Justificativa e importância.....	2
1.3 Objetivos	3
1.3.1 Objetivo geral.....	3
1.3.2 Objetivos específicos.....	4
1.4 Estrutura da dissertação.....	4
2. Características da área e das espécies estudadas	5
2.1 Delineamento experimental.....	5
2.2 Características da área de estudo	6
2.3 Características das espécies estudadas.....	7
2.3.1 <i>Carapa guianensis</i>	7
2.3.2 <i>Cecropia sciadophylla</i>	8
2.3.3 <i>Guarea carinata</i>	8
2.3.4 <i>Guatteria poeppigiana</i>	9
2.3.5 <i>Jacaranda copaia</i>	9
2.3.6 <i>Talisia macrophylla</i>	10

3. A Técnica utilizada	11
3.1 Modelos Lineares Generalizados.....	11
3.1.1 Introdução	11
3.1.2 Distribuição binomial	12
3.1.3 Ligações canônicas.....	13
3.1.3.1 Ligações canônicas para o modelo binomial	14
3.1.4 Função desvio	16
3.1.5 Técnicas de diagnóstico.....	16
3.1.5.1 Gráfico normal de probabilidade para r_i^D com envelope.....	17
3.1.5.2 Gráfico de r_i^D versus valores ajustados $\hat{\mu}_i$	17
3.1.5.3 Gráfico de índices	17
3.1.5.4 Seleção de modelos.....	18
3.1.6 Regressão Logística.....	18
3.2 Redes Neurais Artificiais	19
3.2.1 Introdução	19
3.2.2 Neurônio e funções de ativação de uma Rede Neural Artificial	20
3.2.3 Treinamento de uma Rede Neural Artificial	20
3.2.4 Rede Neural Artificial multicamadas.....	21
3.3 Métricas para avaliar a classificação	22
3.4 Amostragem para validação	23
3.5 A metodologia utilizada para obtenção dos indicadores	24
3.5.1 A estrutura do banco de dados	24
3.6 Tratamentos silviculturais	27
4. Resultados	28
4.1 Tratamento estatístico	28
4.2 Resultados e discussões.....	28
4.2.1 Estatísticas descritivas	28
4.2.2 Modelagem da mortalidade por meio da regressão binária.....	36
4.2.3 Qualidade do ajuste do modelo binomial com suas Funções de ligação	38

4.2.4 Treinamento e validação do modelo	42
4.3 Avaliação da Rede Neural Artificial.....	43
5. Considerações finais e recomendações para trabalhos futuros	45
5.1 Cosiderações finais	45
5.2 Recomendações	46
Referências bibliográficas	47

LISTA DE TABELAS

Tabela 4.1 - Estatísticas descritivas das variáveis independentes quantitativas para as 26.313 unidades amostrais	29
Tabela 4.2 - Espécies selecionadas em função do balanceamento dicotômico.....	31
Tabela 4.3 – Estatística descritiva das variáveis selecionadas para o modelo logístico final.....	31
Tabela 4.4 – Classes comparativas DAP_{ifinal} (cm) para a condição de mortalidade das espécies estudadas.....	33
Tabela 4.5 – Quantidade de unidades amostrais vivas e mortas por tratamento.....	36
Tabela 4.6 – Modelo ajustado para os dados de mortalidade da regressão logística múltipla, com as estimativas, erro, Z e P, considerando a função de ligação logit.....	40
Tabela 4.7 – Taxas de acerto para a regressão logística e RNA das variáveis significativas do modelo, tanto para treinamento quanto para validação	44

LISTA DE FIGURAS

Figura 1.1- Desenho esquemático de uma parcela permanente de 0,25ha.....	6
Figura 3.1 - Gráfico comparativo entre as funções de ligação	15
Figura 3.2 – Quadro da matriz de confundimento.....	22
Figura 4.1 - Histograma do diâmetro altura do peito (DAP_{ifinal} (cm)).....	29
Figura 4.2 - Boxplot do diâmetro altura do peito (DAP_{ifinal} (cm)) por espécies	32
Figura 4.3 - Boxplot da área basal da parcela (G_{pfinal} (cm ²)) por espécies	34
Figura 4.4 – Boxplot das unidades amostrais na subparcela ($P_{spfinal}$) versus código das espécies	35
Figura 4.5 - Boxplot do índice de Shannon (H_{pnorm}) por espécies	35
Figura 4.6 - Diagnóstico para um modelo ajustado a dados com com função de ligação logística (logit)	41
Figura 4.7 - Gráfico normal de probabilidade para um modelo ajustado a dados com função de ligação logística (logit)	42

LISTA DE ABREVIATURAS E SIGLAS

DAP_{ij} (cm) – diâmetro a altura do peito da unidade amostral i no ano de medição j .

$DAP_{i\text{inicial}}$ (cm) – valor inicial do diâmetro a altura do peito da unidade amostral i .

$DAP_{i\text{final}}$ (cm) – valor final do diâmetro a altura do peito da unidade amostral i .

$DAP_{i\text{médio}}$ (cm) – média do diâmetro a altura do peito da unidade amostral i nos anos de medição.

$G_{.pj}$ (cm²) – área basal da parcela p no ano j .

$G_{.p60j}$ (cm²) – área basal da parcela p das unidades amostrais com DAP acima de 60 cm no ano j .

$G_{.spj}$ (cm²) – área basal da subparcela p no ano j .

$G_{.sp60j}$ (cm²) – área basal da subparcelas p das unidades amostrais com DAP acima de 60 cm no ano j .

$ICA_{ij,j+1}$ – Incremento Corrente Anual do DAP da unidade amostral i referente ao ano de medição $j+1$, considerando o ano de medição j .

$P_{.pj}$ – número de unidades amostrais na parcela p no ano j .

$P_{.p60j}$ – número de unidades amostrais na parcela p com DAP acima de 60 cm no ano j .

$P_{.spj}$ – número de unidades amostrais na subparcelas p no ano j .

$P_{.sp60j}$ – número de unidades amostrais na subparcelas p com DAP acima de 60 cm no ano j .

$H_{.pj}$ – indicador de diversidade de Shannon para parcela p no ano de medição j .

$H_{.p\text{min}}$ – valor mínimo do indicador de diversidade de Shannon para parcela p nos anos de medição.

$H_{.p\text{max}}$ – valor máximo do indicador de diversidade de Shannon para parcela p nos anos de medição.

$H_{p\text{médio}}$ – média do indicador de diversidade de Shannon para parcela p nos anos de medição.

$H_{p\text{norm}}$ – valor normalizado do indicador de diversidade de Shannon para parcela p nos anos de medição.

MMV – Método de máxima verossimilhança e,

OR – Odds Ratios (Razão de chances).

TG – Taxa global de classificação correta.

TS – Taxa de sensibilidade de verdadeiro – positivos.

TE – Taxa de especificidade de verdadeiro – negativos.

Capítulo 1

INTRODUÇÃO

Neste capítulo, na seção 1.1 são apresentados os aspectos gerais, na seção 1.2 mostra a justificativa e importância do trabalho, na seção 1.3 pode-se os objetivos (geral e específico), na seção 1.4 observa-se a estrutura da dissertação.

1.1 Aspectos gerais

O Brasil é o país que apresenta a maior biodiversidade do planeta com mais de 55 mil tipos de espécies diferentes, isto equivale a aproximadamente 25% dos tipos de plantas existentes em todo o planeta (SFB, 2010).

Essa biodiversidade se deve a floresta amazônica, a maior biodiversidade existente no planeta (SFB, 2010). As florestas tropicais possuem um papel importante na regulação climática, conservação do solo e da água e são uma fonte de recursos para produção de madeira, remédios, alimentos, minerais etc. A sua diversidade vegetal está bastante relacionada ao processo de regeneração natural das espécies que pode ser um processo natural ou antrópico (OLIVEIRA *et al.*, 2005).

É necessário dispor de informações e instrumentos que descrevam o estado provável das florestas dadas intervenções alternativas de gestão, a fim de tomar decisões mais bem informadas e políticas de manejo florestal sustentável e equitativo (PHILLIPS, *et al.*, 2003, QUEIROZ, 2012).

Há uma quantidade limitada de estudos a respeito de modelos para predição que auxiliem os gestores florestais a uma extração madeireira sustentável. Na área deste estudo, Floresta Nacional do Tapajós, Phillips e tal (2003) apresentaram uma solução para predição do estado futuro das árvores, de acordo com as intervenções de gestão de tratamentos silviculturais, comumente recomendadas para crescimento, recrutamento e mortalidade das espécies arbóreas das florestas tropicais.

Oliveira *et al* (2005) analisaram o efeito de quatro intensidades de redução da área basal, representadas pela exploração e desbastes sistemáticos na composição florística, diversidade, estrutura e crescimento do extrato arbóreo.

No que tange ao problema da avaliação de crescimento (RAMSEY *et al*, 2012; ÖZÇELİK *et al*, 2013; DIAMANTOPOLOU *et al*, 2015) e para a mortalidade de espécies (KING, BENNETT e LIST, 2000) são utilizadas ferramentas estatísticas e soluções computacionais. Neste trabalho serão utilizados os Modelos Lineares Generalizados e Redes Neurais para descrever a mortalidade de seis espécies de árvores: *Carapa guianensis*, *Cecropia sciadophylla*, *Guarea carinata*, *Guatteria poeppigiana*, *Jacaranda copaia* e *Talisia macrophylla*, a partir de indicadores de competição e diversidade na Floresta Nacional do Tapajós, considerando as informações das medições dos anos 1981, 1983, 1987, 1989, 1995, 2003, 2008 e 2012.

1.2 Justificativa e importância

A predição da mortalidade de espécies é uma difícil tarefa, já que a morte (natural) é um evento raro. A grande diversidade de espécies e diferentes comportamentos dificultam ainda mais a classificação, já que se mostra necessário um modelo individual para cada espécie. Outro fenômeno é a ambiguidade dos dados, que também precisa ser estudada, pois, é possível encontrar espécies que morrem ou vivem em situações similares (KING, BENNETT E LIST, 2000).

A grande diferença nos modelos de mortalidade de árvores está no número de espécies que compõem cada floresta. Em um estudo da mortalidade de árvores na Áustria, Monserud e Sterba (1999) modelaram a mortalidade das seis principais espécies que ocorrem nas florestas desse país.

Na Noruega, Eid e Tuhus (2001) desenvolveram modelos de mortalidade para as três principais espécies de árvores do país e ainda para um grupo com as demais espécies. Em Alberta, no Canadá foram construídos modelos de mortalidade para três espécies de árvores (YAO; TITUS; MACDONALD, 2001) e também para uma única espécie (YANG; TITUS; HUANG, 2003).

No caso das florestas tropicais, o número de espécies é muito superior ao das florestas temperadas do hemisfério norte onde foram desenvolvidos os estudos supracitados. O que de certa forma dificulta a modelagem da mortalidade, como no estudo desenvolvido por Vanclay (1991) no norte de Queensland, na Austrália, para 100 espécies de árvores, em que o autor necessitou agrupar as espécies para a construção de 10 modelos de mortalidade, e mesmo assim, não obteve ótimos modelos.

O uso de modelos de mortalidade em florestas fornece um registro da evolução do conhecimento sobre as funções e estrutura da floresta, bem como favorecem o uso de métodos sofisticados para gerar as predições da mortalidade em longo prazo (ROSSI *et al.*, 2007), assim, dentro deste contexto, uma ferramenta bastante útil e amplamente utilizada em diversas áreas são os modelos estatísticos de regressão, e especificamente os Modelos Lineares Generalizados (MLG).

Estes modelos representam uma extensão dos modelos lineares clássicos e têm sido empregados em estudos diversos, apresentando resultados satisfatórios que justificam sua relativa importância no campo da pesquisa científica. O uso da técnica da Inteligência Computacional na análise e interpretação das informações oriundas dos dados observados, em ambientes de incerteza, serve como subsídio para o desenvolvimento de um processo de previsão mais eficaz (CHAVES, 2009).

1.3 Objetivos

1.3.1 Objetivo geral

Desenvolver uma metodologia para estudar a mortalidade de seis espécies arbóreas: *Carapa guianensis*, *Cecropia sciadophylla*, *Guarea carinata*, *Guatteria poeppigiana*, *Jacaranda copaia* e *Talisia macrophylla*, bem como investigar os fatores, utilizando os Modelos Lineares Generalizados e Redes Neurais Artificiais, que associados a competição e a diversidade podem levar a mortalidade das espécies, na Floresta Nacional do Tapajós em Belterra a partir de indicadores de competição e diversidade considerando as informações das medições dos anos 1981, 1983, 1987, 1989, 1995, 2003, 2008 e 2012.

1.3.2 Objetivos específicos

- a) Identificar por meio de Modelos Lineares Generalizados e Redes Neurais Artificiais fatores relacionados à mortalidade das seis espécies arbóreas;
- b) Avaliar os Modelos Lineares Generalizados binomial com diferentes funções de ligação para descrever com confiabilidade estatística a predição da mortalidade das seis espécies arbóreas;
- c) Implementar a Rede Neural Artificial para a verificação correta de mortalidade das unidades amostrais.

1.4 Estrutura da dissertação

Capítulo 1: Refere-se à introdução do trabalho, onde são abordados os aspectos gerais, justificativa e importância do trabalho, objetivos (geral e específico), bem como a estrutura da dissertação;

Capítulo 2: Apresenta uma abordagem sobre a caracterização da área e das seis espécies estudadas;

Capítulo 3: Segue a revisão abordando os Modelos Lineares Generalizados, sua estrutura, alguns parâmetros que permitem verificar a qualidade do ajuste, bem como aborda especificamente o modelo binomial com as funções de ligação, a regressão logística e as Redes Neurais Artificiais;

Capítulo 4: Apresenta os resultados onde é feito o ajuste dos modelos aos dados das seis espécies considerando o melhor ajuste e implementa a RNA;

Capítulo 5: Descreve as considerações finais e recomendações para trabalhos futuros.

Capítulo 2

CARACTERÍSTICAS DA ÁREA E DAS ESPÉCIES ESTUDADAS

Este capítulo apresenta a estrutura do inventário da floresta nacional do tapajós, assim como características específicas da área de estudo. Neste sentido, a Seção 2.1 apresenta o delineamento experimental, a Seção 2.2 aborda as características da área de estudo, a Seção 2.3 trata das características das espécies estudadas.

2.1 Delineamento experimental

Em 1981 a equipe de manejo florestal da Empresa Brasileira de Pesquisas Agropecuária (EMBRAPA), deu início ao experimento e as atividades de inventário florestal por amostragem para determinação da área basal na Floresta Nacional do Tapajós, utilizando o delineamento estatístico de blocos ao acaso com quatro repetições. Cada bloco possui uma área de 36ha, sendo 9ha por tratamento, em cada tratamento foram instaladas ao acaso três parcelas permanentes de 0,25ha, totalizando 12 parcelas por bloco, perfazendo um total de 48 parcelas. A testemunha ou controle do experimento foi estabelecida em 36ha de floresta não explorada, onde foram instaladas 12 parcelas de 0,25ha. No total o experimento possui 60 parcelas permanentes de 0,25ha (OLIVEIRA, 2005).

Com base nas 60 parcelas permanentes o estudo foi desenvolvido com monitoramento contínuo da floresta ao longo do tempo. As medições se deram em árvores com diâmetro a altura do peito (DAP) igual ou superior a 5,0cm em anos não equidistantes. Para facilitar o monitoramento, estas parcelas foram subdivididas em 25 unidades de observações de 10m x 10m e foram numeradas de 1 a 25, iniciando sempre pelo canto sudoeste, conforme ilustra a Figura 1.1. Após a primeira medição realizada pela EMBRAPA foi preparada uma listagem contendo as espécies identificadas pelo seu nome comum e pelos identificadores botânicos.

Figura 1.1: Desenho esquemático de uma parcela permanente de 0,25ha.

5	6	15	16	25
4	7	14	17	24
3	8	13	18	23
2	9	12	19	22
1	10	11	20	21

Fonte: Adaptado de Oliveira (2005).

2.2 Características da área de estudo

O experimento está localizado na Floresta Nacional do Tapajós, município de Belterra, Pará, a altura do km 114 da Rodovia Santarém-Cuiabá entre as coordenadas 2° 40' – 4° 10' de Latitude Sul e 54° 45' - 55° 30' de Longitude Oeste. A altitude está em torno de 175 m acima do nível do mar sendo o relevo plano e levemente ondulado.

O solo predominante é o latossolo amarelo moderado com textura pesada (60% a 94% de argila), com inclusão de latossolo amarelo concrecionário, derivado de argila pedregosa (CARVALHO, 1992).

O clima é do tipo Ami, segundo a classificação de Koppen, que é um clima tropical com uma estação seca de 2 a 3 meses por ano e precipitação anual acima de 2000 mm, com um período chuvoso de março a maio e período seco ou pouco chuvoso de agosto a novembro. A umidade relativa é de aproximadamente 86% (76 a 93%) e a média anual de temperatura está em torno de 25°C, com médias mínimas de 18,4°C e máximas de 32,6° C (CARVALHO, 1992, 2001; OLIVEIRA, 2005).

2.3 Características das espécies estudadas

Para este estudo foram escolhidas seis espécies arbóreas que apresentaram quantidades balanceadas de unidades amostrais vivas e mortas por causas naturais: *Carapa guianensis*, *Cecropia sciadophylla*, *Guarea carinata*, *Guatteria poeppigiana*, *Jacaranda copaia* e *Talisia macrophylla*.

2.3.1 *Carapa guianensis*

A *Carapa guianensis* conhecida como andiroba pertence à família Meliaceae é uma árvore de porte grande com crescimento rápido, atingindo até 30 metros de altura, copa densa e geralmente com sapopemas baixas. É amplamente distribuída em toda Amazônia e ocorre em alta frequência em floresta de terra firme e na várzea. Na várzea, atinge uma altura maior e ocorre em maior densidade (EMBRAPA, 2004).

Sua madeira é utilizada na construção civil (vigas, caibros, ripas, esquadrias de portas, lambris, venezianas, batentes, caixilhos, rodapés, guarnições, molduras), móveis, lâminas, compensados, caixas de embalagem, acabamentos internos de barcos e navios e outros (MAUÉS, 2006).

A espécie tem grande valor comercial, o óleo das sementes é um dos produtos medicinais mais vendidos. Na medicina popular, é usado para aliviar contusões edemas, reumatismo, para auxiliar na cicatrização e recuperar a pele. O chá da casca é muito amargo e é utilizado na cura da febre e também muito utilizado como repelente de insetos e na fabricação de vela, sabão e xampu. A madeira é resistente contra ataque de insetos. A casca transformada em pó serve para tratar feridas e vermes (GONÇALVES, 2001).

Alguns indivíduos podem ter flores em qualquer mês, porém o pico de floração é entre agosto e novembro. Os frutos precisam de quase um ano para amadurecer, a dispersão ocorre principalmente entre abril e junho (EMBRAPA, 2004).

2.3.2 *Cecropia sciadophylla*

A *Cecropia sciadophylla* é conhecida como embaúba-torém. Embaúba significa fruta da árvore de tronco oco em tupi guarani porque algumas espécies de formigas vivem dentro do tronco (PIO CORRÊA, 1978; BERG; ROSSELLI, 2005).

O *Cecropia* é um dos maiores gêneros pioneiros de árvore com 61 espécies. Este gênero é encontrado nas formações secundárias e nas clareiras no interior das florestas, diferencia-se dos demais gêneros de Urticaceae por possuir folhas palmatilobadas e inflorescências protegidas por espata (BERG; ROSSELLI, 2005; GAGLIOTI, 2011).

A *Cecropia sciadophylla* pode ser diferenciada das demais espécies pelas incisões profundas no limbo que vão até o pecíolo e pela ausência de triquílios (conjunto de tricomas na base do pecíolo). A espécie floresce durante longo período do ano, predominando em agosto-novembro. Os frutos amadurecem em novembro-fevereiro (BERG; ROSSELLI, 2005). É uma das maiores Cecrópias, tendo a árvore de 10 m a 35 m de altura. Tronco reto e roliço, oco, com casca lisa marcada pelas cicatrizes das estípulas e pecíolos antigos (SILVA, 2012).

A espécie é comumente encontrada em clareiras de florestas maduras. Sua ocorrência na Amazônia se dá nos estados do Amapá, Pará, Roraima, Amazonas, Rondônia e Acre. Sua madeira é empregada para caixotaria, forros, palitos de fósforo, pasta celulósica e balsa para transporte de madeira pesada nos rios amazônicos. Os frutos são muito procurados por aves e outros animais e as folhas são o alimento principal do bicho-preguiça (LORENZI, 2002).

2.3.3 *Guarea carinata*

O gênero *Guarea* pertence à família Meliaceae com cerca de 60 espécies e está incluído entre os seis gêneros da família mais encontrados na região amazônica (PEREIRA, WOLTER, ROCHA, CARVALHO, BRAZ, 1992).

A *Guarea carinata* popularmente conhecida como andirobarana vermelha, tem maior incidência nos estados do Pará, Roraima, Acre, Amazonas, Rondônia e Mato Grosso. Árvore de 12 a 15 metros de altura, o tronco varia de 25 a 30 cm de diâmetro (PEREIRA, WOLTER, ROCHA, CARVALHO, BRAZ, 1992).

2.3.4 *Guatteria poeppigiana*

A *Guatteria poeppigiana* popularmente conhecida como envira preta pertence a família Annonaceae é uma árvore de grande porte que pode atingir 22 metros de altura. Madeira amarelada leve com flores esverdeadas, sendo sua madeira utilizada na confecção de cabo para rodo (PORTAL DA BIODIVERSIDADE).

Tem grande ocorrência no norte do Brasil nos Estados do Pará, Amazonas, Amapá não sendo uma espécie ameaçada (FLORA DO BRASIL).

2.3.5 *Jacaranda copaia*

A *Jacaranda copaia* popularmente conhecida como parapará pertence a família Bignoniaceae é uma árvore de porte grande, atingindo de altura de 20 a 35 metros de altura, com ocorrência na região amazônica (GENTRY; MORAWETZ, 1992).

O gênero *Jacaranda* têm cerca de 50 espécies, todas neotropicais. A maioria delas é encontrada principalmente nos cerrados e outros ambientes mais secos ao redor da Amazônia. A *Jacaranda copaia* é a única espécie deste gênero amplamente distribuída na Amazônia (EMBRAPA, 2004).

A árvore cresce muito rapidamente, o que é importante para a regeneração de matas de capoeira. Por isso, esta espécie é bastante indicada para uso em reflorestamento. Quando floresce, é uma linda árvore, ótima para arborização urbana. A base nunca tem sapopemas altas ou planas. Pode ser um pouco dilatada ou ter raízes grossas elevadas (LORENZI, 1992).

A *Jacaranda copaia* apresenta uma madeira muito leve, inadequada para movelaria e indicada para lenha e construção leve, por exemplo: brinquedos, caixas, cabo de faca, etc. Madeira com cerne e alburno indistintos, de cor branco-palha a levemente rosada, apresentando listras vasculares mais escuras; grã direita, textura grosseira; superfície lustrosa podendo receber bom acabamento (RIBEIRO, *et al.*, 1999).

Na Amazônia a floração ocorre principalmente entre agosto e setembro, mas algumas árvores florescem fora desta época. Os frutos secos podem continuar na árvore após a dispersão das sementes (EMBRAPA, 2004).

2.3.6 *Talisia macrophylla*

A *Talisia macrophylla* popularmente conhecida como pitombarana, pitomba, pitomba-do-mato e pitomba folha grande, pertence a família do gênero Sapindaceae a espécie que pode atingir até 12 metros de altura. Segundo Rodrigues (2003) a *Talisia Macrophylla* tem de ocorrência restrita ao domínio amazônico.

A *Talisia macrophylla* é um tipo de palma de pequeno a médio porte, as florestas de árvores desta espécie tem grandes folhas verdes brilhantes e grandes aglomerados de frutas amarelas escuras (ICMbio, 2017).

Sua distribuição geográfica está ligada aos estados da região Norte como: Pará, Amapá, Roraima, Amazonas, Rondônia e nordeste pelo estado da Bahia. A *Talisia macrophylla* ocorre no domínio Atlântico, tratando-se de espécie relativamente comum no sub-bosque das florestas ombrófilas, suas flores apresentam cinco estames, coletada com flores em setembro (RODRIGUES, 2003).

Capítulo 3

A TÉCNICA UTILIZADA

O capítulo descreve na Seção 3.1 a introdução a respeito dos modelos lineares generalizados e as características do modelo de distribuição binomial; a Seção 3.2 fornece um conjunto de informações referente a técnica de Redes Neurais Artificiais e a Seção 3.3 mostra as métricas para avaliar a classificação, na seção 3.4 tem-se a amostragem para validação, na seção 3.5 e mostrada a metodologia utilizada para obtenção dos indicadores e a seção 3.6 os tratamentos silviculturais adotados.

3.1 Modelos Lineares Generalizados

3.1.1 Introdução

O modelo de regressão normal linear atendeu aos objetivos pretendidos nessas análises por muitas décadas. Entretanto, surgiram outras situações que não puderam ser explicadas adequadamente pelo fato da variável resposta seguir uma distribuição não normal, ou ainda, por possuir um caráter não linear. Outros modelos surgiram na tentativa de explicar essas situações tais como o modelo *probit*, os modelos *log-lineares*, o modelo complementar *log-log* e outros. Em 1972, Nelder e Wedderburn propuseram os Modelos Lineares Generalizados (MLGs) em que as opções para a distribuição da variável resposta são estendidas, desde que estas sejam integrantes de famílias exponenciais (PAULA, 2010; CORDEIRO e DEMÉTRIO, 2013).

A família exponencial é caracterizada por uma função (de probabilidade ou densidade) considerando y_1, \dots, y_n variáveis aleatórias independentes, cada uma com função densidade (f.d.p) ou função de probabilidades (f.p) na forma (3.1) especificada para qualquer i (PAULA, 2010; CORDEIRO e DEMÉTRIO, 2013).

$$f(y_i; \theta_i; \phi) = \exp[\phi\{y_i\theta_i - b(\theta_i)\} + c(y_i, \phi)] \quad (3.1)$$

onde $E(Y_i) = \mu_i = b'(\theta_i)$ e $Var(Y_i) = \phi^{-1}V(\mu_i)$, em que $V_i = V(\mu_i) = \frac{d\mu_i}{d\theta_i}$ é a função de variância e $\phi^{-1} > 0$ é o parâmetro de dispersão (PAULA, 2010; CORDEIRO e DEMÉTRIO, 2013).

Nesses modelos, a média da variável resposta y é modelada por meio de uma estrutura de regressão que envolve parâmetros desconhecidos, covariáveis e uma função de ligação (3.2).

$$g(\mu_i) = \eta_i \quad (3.2)$$

em que $\eta_i = X_i^T \beta$ é o preditor linear, $\beta = (\beta_1, \dots, \beta_p)^T$, $p < n$, é um vetor de parâmetros desconhecidos a serem estimados, $X_i = (x_{i1}, \dots, x_{ip})^T$ representa os valores de variáveis explicativas e $g(\cdot)$ é uma função monótona e diferenciável, denominada função de ligação (PAULA, 2010).

3.1.2 Distribuição binomial

Os modelos de regressão para dados binários são utilizados quando a variável dependente do modelo assume apenas valores zero ou um. Esses modelos são amplamente utilizados em diversas áreas de aplicação (McCULLAGH e NELDER, 1989; AYALEW e YAMAGISHI, 2005).

Se Y é a proporção de sucessos em n ensaios independentes, cada um com probabilidade de ocorrência μ . Assumindo que $nY \sim B(n, \mu)$. A função de probabilidades de Y fica então expressa na forma (3.3).

$$\binom{n}{ny} \mu^{ny} (1 - \mu)^{n-ny} = \exp \left\{ \log \binom{n}{ny} + ny \log \left(\frac{\mu}{1-\mu} \right) + n \log(1 - \mu) \right\} \quad (3.3)$$

em que $0 < \mu, y < 1$. Obtemos (3.1) fazendo $\phi = n$, $\theta = \log \left\{ \frac{\mu}{1-\mu} \right\}$, $b(\theta) = \log(1 + e^\theta)$ e $c(y, \phi) = \log \binom{\phi}{\phi y}$. A função de variância é dada por $V(\mu) = \mu(1 - \mu)$.

A relação entre a média do componente aleatório, y , e o preditor linear do modelo é apresentada por meio de uma função de ligação. Para cada distribuição para a variável dependente em um MLG existem diferentes funções de ligação que podem ser utilizadas (McCULLAGH e NELDER, 1989). Quando as inferências do modelo são feitas via máxima verossimilhança, como nos MLG's, é usual a utilização das seguintes estatísticas para testar a significâncias de covariáveis no modelo: razão de verossimilhanças, Wald e escore (PAULA, 2010; CORDEIRO e DEMÉTRIO, 2013).

3.1.3 Ligações canônicas

Se ϕ é conhecido, o logaritmo da função de verossimilhança de um modelo linear generalizado com variáveis respostas independentes pode ser expresso da por (3.4).

$$L(\beta) = \sum_{i=1}^n \phi \{y_i \theta_i \quad b(\theta_i)\} + \sum_{i=1}^n c(y_i, \phi) \quad (3.4)$$

Em particular se tem um caso importante quando ocorre que o parâmetro canônico (θ) coincide com o preditor linear, ou seja, $\theta_i = \eta_i = \sum_{j=1}^p x_{ij} \beta_j$. Nesse caso, $L(\beta)$ se torna (3.5) fica:

$$L(\beta) = \sum_{j=1}^p S_j \beta_j \quad \phi \sum_{i=1}^n b \left(\sum_{j=1}^p x_{ij} \beta_j \right) + \sum_{i=1}^n c(y_i, \phi) \quad (3.5)$$

Com $S_j = \phi \sum_{i=1}^n Y_i x_{ij}$.

A estatística $S = (S_1, \dots, S_p)^T$ é pelo teorema da fatoração suficiente mínima para o vetor $\beta = (\beta_1, \dots, \beta_p)^T$. As ligações correspondentes a determinadas estatísticas são denominadas ligações canônicas têm um importante papel para os modelos lineares generalizados (MLGs), (WEDDERBURN, 1976).

3.1.3.1 Ligações canônicas para o modelo binomial

A seleção da função de ligação nos MLG's é muito importante uma vez que a utilização de uma função de ligação inadequada pode levar a erro do cálculo do componente de desvio (deviance) e também ter inferências incorretas sobre os β 's, acarretando uma interpretação errada do modelo (McCULLAGH e NELDER, 1989).

As ligações canônicas mais comuns para o modelo binomial são as ligações logística, probito e complemento log-log. Neste trabalho foram testados estas três funções de ligação e como não houve grandes mudanças nos parâmetros estimados e optou-se por trabalhar com a função de ligação logística que será abordada na seção 3.1.6.

a) Ligação logística (logit)

A função logit para o modelo binomial tem função de ligação canônica, dada por (3.6):

$$\log \left\{ \frac{\mu}{1-\mu} \right\} = \eta \quad (3.6)$$

b) Ligação probito (probit)

Seja μ e a proporção de sucessos de uma distribuição binomial. A função de ligação probito é definida por (3.7):

$$\Phi^{-1}(\mu) = \eta \quad (3.7)$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada da normal padrão.

c) Ligação complemento log-log (cloglog)

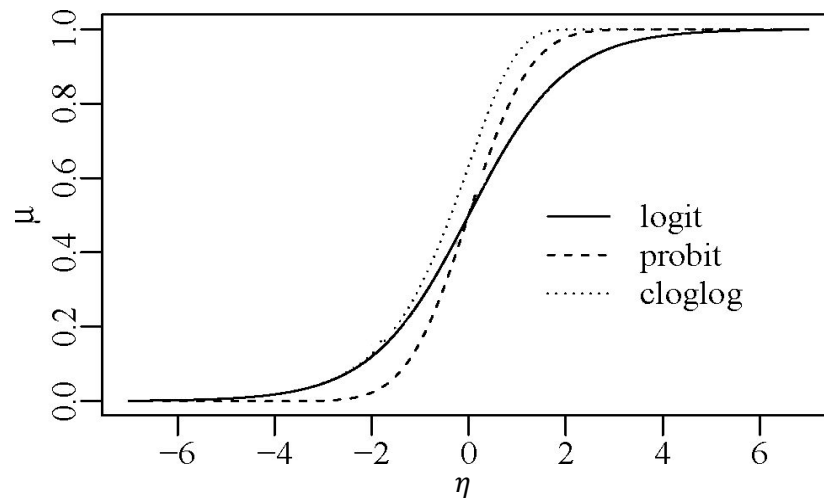
O modelo binomial com ligação complemento log-log é definido por (3.8) ou de forma similar por (3.9):

$$\mu = 1 - \exp \{ -\exp(\eta) \} \quad (3.8)$$

$$\log\left\{\frac{\mu}{1-\mu}\right\} = \eta \quad (3.9)$$

A Figura 3.1 apresenta uma comparação gráfica das funções de ligação consideradas. Pode-se verificar que a função de ligação cloglog é assimétrica, diferentemente das outras duas funções. Para $\eta \approx 0$, as funções de ligação logit e probit levam a valores muito próximos de μ . Quando $\eta \approx \pm 1,5$ as funções de ligação logit e cloglog são muito semelhantes. É possível observar que as três funções são muito similares quando valores de μ são próximos dos extremos do intervalo padrão (0,1) (CANTERLE, BAYER, 2015).

Figura 3.1: Gráfico comparativo entre as funções de ligação.



Fonte: Extraído de CANTERLE, BAYER (2015).

3.1.4. Função desvio

O logaritmo da função de verossimilhança será definido por (3.10):

$$L(\mu; y) = \sum_{i=1}^n L(\mu_i; y_i) \quad (3.10)$$

em que $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = X_i^T \beta$. Para o modelo saturado ($p = n$) a função $L(\mu; y)$ é estimada por (3.11):

$$L(y; y) = \sum_{i=1}^n L(y_i; y_i) \quad (3.11)$$

A estimativa de máxima verossimilhança de μ_i fica dada por $\tilde{\mu}_i = y_i$. Quando $p < n$, denota-se a estimativa de $L(\mu; y)$ por $L(\mu; y)$. Neste caso, a estimativa de máxima verossimilhança de μ_i será dada por $\mu_i = g^{-1}(\eta_i)$, em que $\eta_i = X_i^T \hat{\beta}$.

A qualidade do ajuste de um modelo linear generalizado (MLG) é avaliada por meio da função desvio (3.12).

$$D(y; \mu) = \phi D(y; \mu) = 2\{L(y; y) - L(\mu; y)\} \quad (3.12)$$

A função desvio é uma distância entre o logaritmo da função de verossimilhança do modelo saturado (com n parâmetros) e do modelo sob investigação (com p parâmetros) avaliado na estimativa de máxima verossimilhança $\hat{\beta}$ (PAULA, 2010).

Um valor pequeno para a função desvio indica que, para um número menor de parâmetros, se obtém um ajuste tão bom quanto o ajuste com o modelo saturado. A função desvio da distribuição binomial é definida por (3.13) (PAULA, 2010).

$$D(y; \mu) = 2 \sum_{i=1}^k y_i \log \left(\frac{y_i}{n_i \hat{\mu}_i} \right) + (n_i - y_i) \log \left\{ \frac{\left(\frac{1-y_i}{n_i} \right)}{(1-\hat{\mu}_i)} \right\} \quad (3.13)$$

3.1.5 Técnicas de diagnóstico

As técnicas de diagnóstico utilizadas neste trabalho foram análise visual de gráficos para detectar padrões ou pontos atípicos. Para a geração destes gráficos utilizou-se o software R baseados em Paula (2010).

3.1.5.1 Gráfico normal de probabilidade para r_i^D com envelope

O gráfico normal de probabilidade é um recurso permite detectar o afastamento da suposição de normalidade dos resíduos. Espera-se ainda, que ausência de pontos discordantes, seu aspecto seja o de uma reta, sendo colinear da reta ajustada. Entretanto, percebe-se que dificilmente os pontos do gráfico seguem exatamente a disposição da reta ajustada mantendo certa variação em torno desta. E em virtude da disposição irregular dos resíduos criou-se uma de banda de confiança a partir dos resíduos obtidos do modelo ajustado por meio de simulações, esta banda é denominada de envelope (PAULA, 2010; OLIVEIRA, 2013).

3.1.5.2 Gráfico de r_i^D versus valores ajustados $\hat{\mu}_i$

A heterocedasticidade da variância e a presença de pontos discrepantes são visualizadas neste gráfico. Segundo o que se espera é que o gráfico apresente a distribuição dos resíduos em torno de zero com amplitude constante, onde desvios sistemáticos podem ter algum tipo de curvatura ou uma amplitude muito diferente do valor ajustado (MARCIANO, 2009).

3.1.5.3 Gráfico de índices

Este tipo de gráfico é gerado assumindo-se que no eixo horizontal estão as ordens das observações, no qual é possível verificar a presença de pontos atípicos que poderão ser classificados como: influentes, não influentes ou como pontos aberrantes (*outlier*). Para o presente trabalho destacam-se os seguintes gráficos

- i. *deviance* residual padronizada (r_i^D): útil para detectar pontos aberrantes;
- ii. pontos de alavanca ii (*leverage*): utilizado para observar pontos influentes;
- iii. distância de Cook (D_i): usado para confirmar se os pontos tidos como *leverage* são realmente influentes.

3.1.5.4 Seleção de modelos

O método para seleção de modelos utilizado neste trabalho foi o método de Akaike. Este utiliza o logaritmo da função de verossimilhança $L(\beta)$ que cresce com o aumento do número p de parâmetros do modelo, assim o melhor modelo é aquele com menor valor para a função AIC (3.14) (AKAIKE, 1974).

$$AIC = L(\hat{\beta}) + p \quad (3.14)$$

3.1.6 Regressão logística

O modelo de regressão logística pertence à classe dos Modelos Lineares Generalizados com a função de ligação logística. Esta técnica de regressão logística foi desenvolvida por volta da década de 1960 para realizar predições de fenômeno sem que a variável dependente fosse de natureza binária embora possa ser aplicada a dados multinominais, sendo codificada como um número inteiro, variando de 1 a $k - 1$, onde k é um número positivo (HOSMER; LEMESHOW, 1989).

Neste trabalho, esse modelo será ajustado para prever a probabilidade de ocorrência da morte das unidades amostrais, ou seja, $\pi = P[Y = 1]$, a partir do conhecimento das variáveis independentes X_p na equação (3.15).

$$\pi = P[Y = 1] = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}} \quad (3.15)$$

Com os parâmetros do modelo β_0, \dots, β_p estimados pelo método de Máxima Verossimilhança (MMV) com a probabilidade estimada de ocorrência da morte para unidade amostral i obtida por (3.16).

$$\hat{\pi}_i = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_p X_{ip}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_p X_{ip}}} \quad (3.16)$$

A função exponencial dada por (3.17) define o quociente entre a probabilidade de ocorrência e a probabilidade de não ocorrência da morte, dado os valores das variáveis preditoras fixados. Esta que é denominada de chance (HOSMER; LEMESHOW, 1989).

$$c\ chance(X_i) = \frac{\pi(X_i)}{1-\pi(X_i)} = e^{\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_p X_{ip}} \quad (3.17)$$

Muitas vezes é de interesse estimar a contribuição na chance quando se aumenta uma unidade em uma determinada variável preditora, mantendo as demais fixas. Considerando que a i -ésima unidade amostral tem vetor de preditoras $X_i=(X_{i1}, \dots, X_{ik}, \dots, X_{ip})$ e a j -ésima unidade amostral tem vetor de preditoras $X_j=(X_{j1}, \dots, X_{jk}, \dots, X_{jp})$ que difere de X_i apenas na k -ésima preditora e que $X_{jk}=X_{ik}+1$. Neste caso, define-se a *razão de chances* ou odds-ratio (OR) como (3.18).

$$\widehat{OR} = \frac{\text{chance}(X_j)}{\text{chance}(X_i)} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_k X_{jk} + \dots + \hat{\beta}_p X_{ip}}}{e^{\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_k X_{ik} + \dots + \hat{\beta}_p X_{ip}}} = e^{\hat{\beta}_k} \quad (3.18)$$

A estimativa (3.18) representa a contribuição na chance devido ao aumento na preditora X_k . Assim, valores de \widehat{OR} muito próximos de um indica pouca contribuição significativa de X_k ; valores maiores que um indicam aumento da probabilidade de ocorrência, e valores menores que um indicam a diminuição da probabilidade de ocorrência da característica (DIAS FILHO, CORRA, 2007).

3.2 Redes Neurais Artificiais

3.2.1 Introdução

As Redes Neurais Artificiais (RNAs) são técnicas computacionais que apresentam um modelo inspirado na estrutura neural de organismos inteligentes utilizadas nas soluções de problemas (NEGNEVITSKY, 2005).

Neste trabalho estas técnicas foram utilizadas para classificar as unidades amostrais das seis espécies estudadas considerando os indicadores de diversidade e competição.

3.2.2 Neurônio e funções de ativação de uma Rede Neural Artificial

As RNAs consistem em um conjunto de unidades básicas de processamento (neurônios) ligadas entre si por canais de comunicação que estão associados a determinado peso numérico e que transmitem sinais de um neurônio para outro. Estes pesos sinápticos representam a importância de cada entrada do neurônio. Cada neurônio recebe uma quantidade de sinais de entrada e sempre produz um único sinal de saída que pode ser transmitido para outros neurônios (NEGNEVITSKY, 2005; CAVALCANTE, 2013).

A saída de cada neurônio é determinada pela função de ativação a ele associado e pelas conexões por meio das quais recebe entradas e transmite sua saída, portanto, o caminho pelo qual a informação é conduzida. As funções mais utilizadas são as gaussianas, sigmóides, senoidais, lineares, tangentes hiperbólicas, logarítmicas, degrau, entre outras (NEGNEVITSKY, 2005; PALMA NETO e NICOLLETTI, 2005).

3.2.3 Treinamento de uma Rede Neural Artificial

O aprendizado das RNAs ocorre por meio dos ajustes nos pesos de todos os canais da rede. Estes ajustes são definidos pelo algoritmo de treinamento baseado na arquitetura da RNA (CAVALCANTE, 2013). Uma Rede Neural Artificial sempre aprende a partir de um conjunto de treinamento; o aprendizado se repete na capacidade de generalização que a Rede Neural vai exibir, quando usada em novas situações os algoritmos de treinamento da Rede Neural Artificial pertencem a dois grupos: supervisionados e não supervisionados (PALMA NETO e NICOLLETTI, 2005; CHAVES, 2009).

a) **supervisionado**: a descrição de cada exemplo de treinamento traz também a classe que o exemplo representa, assim durante a fase de treinamento, o algoritmo tenta ajustar os pesos sinápticos de maneira que a saída da Rede Neural coincida com a classe associada ao exemplo, para cada exemplo do conjunto de treinamento;

b) **não supervisionado**: a Rede Neural é treinada sem considerar a informação da classe associada a cada exemplo de treinamento e o aprendizado não supervisionado se dá por meio da modificação dos pesos sinápticos de modo que exemplos semelhantes sejam associados uma mesma unidade de saída.

3.2.4 Rede Neural Artificial Multicamada

Nos últimos anos, vários paradigmas neurais foram introduzidos, com destaque para o Perceptron Multicamada (MultilayerPerceptron). Este modelo de RNA consiste em vários neurônios interligados entre si com uma ou mais camadas escondidas (NEGNEVITSKY, 2005).

O algoritmo Backpropagation e suas derivações é um dos algoritmos para treinamento de RNAs multicamadas, mais difundido na comunidade científica. Baseia-se no aprendizado supervisionado por correção de erros. Sua utilização compreende duas fases de propagação:

1º - Propagação: Um padrão é apresentado à camada da entrada da rede. A atividade resultante flui camada por camada até que a resposta seja produzida pela camada de saída, onde é obtida a resposta da rede e o erro é calculado.

2º - Retropropagação ("backpropagation"): são feitas alterações nos pesos sinápticos desde a camada de saída até a camada de entrada.

A RNA pode ser utilizada como uma ferramenta de classificação de novos dados depois que estiver treinada com as novas entradas apresentadas à camada de entrada, as quais são processadas nas camadas intermediárias e os resultados são apresentados na camada de saída, mas sem a retropropagação do erro (CAVALCANTE, 2013).

3.3 Métricas para avaliar a classificação

Neste trabalho, a variável dependente denominada “estado” que representa a condição da unidade quanto à sobrevivência assume dois valores mutuamente exclusivos, viva ou morta, sendo o estado da unidade amostral morto, a classe positiva, enquanto que as unidades amostrais vivas representam a classe negativa. O desempenho do modelo de classificação pode ser medido pela matriz de confundimento (Quadro 3.2).

Figura 3.2: Quadro da matriz de confundimento.

Classes	Predição positiva	Predição negativa
Positiva	Verdadeiro Positivo (VP)	Falso Negativo (FN)
Negativa	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Para avaliar a classificação foram utilizadas as taxas: global de classificação correta, a sensibilidade e a especificidade na amostra em estudo obtidas da matriz de confundimento (CHAVES, 2009).

A taxa global de classificação, (TG), é a razão entre o número de casos que foram classificados corretamente e o número total de casos, e é definida por (3.19):

$$T_G = \frac{VP+VN}{VP+VN+FP+FN} \quad (3.19)$$

A taxa de sensibilidade, (TS), avalia a proporção de verdadeiro-positivos identificados corretamente, e é definida por (3.20).

$$T_S = \frac{VP}{VP+FN} \quad (3.20)$$

E a taxa de especificidade, (TE), avalia a proporção de verdadeiro-negativos identificados corretamente, e é definida por (3.21).

$$T_E = \frac{VN}{VN+FP} \quad (3.21)$$

3.4 Amostragem para validação

Uma base de dados é desbalanceada quando existem menos casos de algumas classes do que de outras. Para evitar esse viés, técnicas de pré-processamento são utilizadas para alterar as distribuições das classes na base de treinamento, visando reduzir o desbalanceamento. Como a eliminação de casos da classe majoritária (*undersampling*) e replicação de casos da classe minoritária (*oversampling*) dos dados de treinamento, visando obter classificadores melhores do que os obtidos a partir da distribuição original (CHAVES, 2009).

Uma abordagem bastante comum e a utilização dos métodos de amostragem, assim para a composição da amostra de validação, utilizou-se a amostragem estratificada para seleção das unidades amostrais para treinamento e validação, o banco de dados composto das seis espécies foi estratificado em k amostras de tamanhos proporcionais a quantidade de vivos e mortos, bem como a das espécies, sendo selecionados aleatoriamente 75% para treinamento e 25% para validação. A amostragem estratificada tem como objetivo reduzir a variância nas estimativas.

Com o intuito de uma melhor classificação somente as variáveis significativas pela Regressão Logística são selecionadas, assim a Rede Neural Artificial é utilizada, pois a inclusão de variáveis não significativas poderiam prejudicar seu desempenho, bem como, levar ao erro de classificação (CHAVES, 2009).

A Regressão Logística foi executada com todas as variáveis definidas na Tabela 4.1. Posteriormente, a Rede Neural Artificial foi executada somente com as variáveis significativas obtidas na regressão logística, considerando o nível de significância $\alpha = 5\%$.

3.5 A metodologia utilizada para obtenção dos indicadores

3.5.1 A Estrutura do banco de dados

O banco de dados original é constituído de uma planilha eletrônica, em que as variáveis mensuradas estavam dispostas em colunas e as unidades amostrais medidas sequencialmente nas linhas por ano, parcela, subparcela, individuo, fuste e código das espécies. Inicialmente foram avaliadas algumas variáveis nominais e ordinais entre as quais: a condição do fuste, se a unidade sofreu algum tipo de dano, se estava podre, o tipo de iluminação, a forma, se tinha cipó e classe de floresta a que pertencia, entretanto, nem todas as informações estavam completas, de modo que se optou a utilizar as variáveis relacionadas com o crescimento e competição.

a) Variável dependente

Com a informação das unidades amostrais mortas foi criada a variável denominada “estado” que representa a condição da unidade quanto à sobrevivência. Esta variável é binária ou dicotômica, definida como: 0 para unidade amostral viva e 1 para morta. No estudo foram consideradas somente unidades amostrais que tiveram a chamada morte natural.

b) Variável independente

As variáveis independentes foram àquelas relacionadas com o crescimento da unidade amostral, bem como, relacionadas à competição. Assim foram consideradas: DAP_{ij} , $G_{\cdot pj}$, $G_{\cdot p60j}$, $G_{\cdot spj}$, $G_{\cdot sp60j}$, $ICA_{ij \cdot j+1}$, $P_{\cdot pj}$, $P_{\cdot p60j}$, $P_{\cdot spj}$, $P_{\cdot sp60j}$ e $H_{\cdot pj}$.

O DAP_{ij} representa o tamanho ou dimensão da unidade amostral i em centímetro no ano de medição j (3.22):

$$DAP_{ij} = \frac{CAP_{ij}}{\pi} \quad (3.22)$$

onde CAP_{ij} é a circunferência à altura do peito medida a 1,3 metros do solo em centímetros e $\pi= 3,142$. Neste trabalho, foram testados os DAP_{ij} médios ($DAP_{imédio}$), bem como os iniciais ($DAP_{inicial}$) e finais (DAP_{ifinal}).

A área basal da parcela ($G_{.pj}$) representa o quanto da superfície da parcela está ocupado por n unidades amostrais. Esta é obtida pelo somatório da área basal de n unidades amostrais na parcela p no ano j é dada por (3.23):

$$G_{.pj} = \sum_{i=1}^n g_{ipj} \quad (3.23)$$

onde a área basal da unidade amostral i no ano de medição j (g_{ij}) em $(cm)^2$ é dada por (3.24):

$$g_{ij} = \frac{\pi \times DAP_{ij}^2}{4} \quad (3.24)$$

A área basal da parcela 60cm ($G_{.p60j}$) representa o quanto da superfície da parcela está ocupado por n unidades amostrais com DAP acima de 60 cm. Esta é obtida pelo somatório da área basal de n unidades amostrais na parcela p no ano j com DAP acima de 60 cm é dada por (3.25):

$$G_{.p60j} = \sum_{i=1}^n g_{ip60j} \quad (3.25)$$

A área basal da subparcela ($G_{.spj}$) representa o quanto da superfície da subparcela está ocupada por n unidades amostrais. Esta é obtida pelo somatório da área basal de n unidades amostrais na subparcelas p no ano j é dada por (3.26):

$$G_{.spj} = \sum_{i=1}^n g_{ispj} \quad (3.26)$$

A área basal da subparcela 60cm ($G_{.sp60j}$) representa o quanto da superfície da subparcela está ocupado por n unidades amostrais com DAP acima de 60 cm. Esta é obtida pela soma da área basal de n unidades amostrais na subparcela p no ano j com DAP acima de 60 cm é dada por (3.27):

$$G_{.sp60j} = \sum_{i=1}^n g_{isp60j} \quad (3.27)$$

O Incremento Corrente Anual do DAP da unidade amostral i referente ao ano de medição $j+1$, considerando o ano de medição j ($ICA_{ij,j+1}$) representa o crescimento em centímetros do diâmetro da unidade amostral em ocorrido entre a quantidade de dias entre duas medições, Δd , é dado por (3.28):

$$ICA_{ij,j+1} = \frac{365,2 \times (DAP_{i,j+1} - DAP_{i,j})}{\Delta d} \quad (3.28)$$

O número de unidades amostrais na parcela p é obtido pela soma de unidades amostrais na parcela p (P_{pj}) no ano j . Já o número de unidades amostrais com DAP acima de 60 cm na parcela p no ano j (P_{p60j}) e é obtido pela soma das unidades amostrais na p no ano j com DAP acima de 60 cm.

O número de unidades amostrais na subparcelas p é obtido pela soma de unidades amostrais na subparcelas p (P_{spj}) no ano j . Já o número de unidades amostrais com DAP acima de 60 cm na subparcela p no ano j (P_{sp60j}) é obtido pela soma das unidades amostrais na subparcelas p no ano j com DAP acima de 60 cm.

O indicador de diversidade de Shannon para parcela p no ano de medição j (H_{pj}) foi utilizado para pra determinar a diversidade de espécies em cada tratamento e nas parcelas e é dado por (3.29). Neste trabalho, este indicador foi normalizado e foram obtidos os indicadores mínimos (H_{pmin}), médios (H_{pmedio}) e máximos (H_{pmax}) das medições.

$$H_{pj} = - \sum_{i=1}^S p_i \times \ln p_i \quad (3.29)$$

onde $p_i = n_i/N$ com n_i sendo o número de indivíduos da espécie i , N representa o número total de indivíduos amostrados e S o número total de espécies na parcela p .

3.6 Tratamentos silviculturais

As florestas são compostas por árvores das mais variadas espécies com tamanhos distintos, variadas e indeterminadas idades e crescendo sobre distintas condições ambientais. Com o objetivo de reduzir a mortalidade das espécies devido a competição existente na floresta foi adotado a técnica de tratamentos silviculturais de blocos ao acaso com quatro níveis de intensidade (OLIVEIRA, 2005).

Tratamentos Silviculturais Adotados:

- a) **T0**: controle, representado pela floresta não explorada;
- b) **T1**: Exploração tradicional – abate de árvores com $DAP \geq 45\text{cm}$, sem nenhuma intervenção posterior. O limite de 45cm de diâmetro foi aplicado por ser o mínimo permitido pela legislação florestal para o corte de árvores em florestas naturais;
- c) **T2**: Abate de árvores comerciais $DAP \geq 55\text{cm}$ + anelamento com desvitalização de espécies não comerciais para reduzir a área basal em 20% da original;
- d) **T3**: Abate de árvores comerciais $DAP \geq 55\text{cm}$ + anelamento com desvitalização de espécies não comerciais para reduzir a área basal em 40% da original;
- e) **T4**: Abate de árvores comerciais $DAP \geq 55\text{cm}$ + anelamento com desvitalização de espécies não comerciais para reduzir a área basal em 60% da original;

Capítulo 4

RESULTADOS

4.1 Tratamento estatístico

Inicialmente os dados foram analisados por meio da estatística descritiva, a fim de se conhecer cada variável. Feito isso, utilizou-se a regressão binária univariada (RBU) e múltipla (RBM) para identificar as variáveis associadas à mortalidade das árvores construindo-se um modelo para explicar a relação entre elas. As análises estatísticas foram realizadas com auxílio do programa R (R PROGRAM, 2014).

4.2 Resultados e discussões

4.2.1. Estatísticas descritivas

O banco de dados originalmente contou com 31.991 unidades amostrais, no entanto considerando as observações que sofreram morte natural este número foi reduzido a 26.313 unidades amostrais.

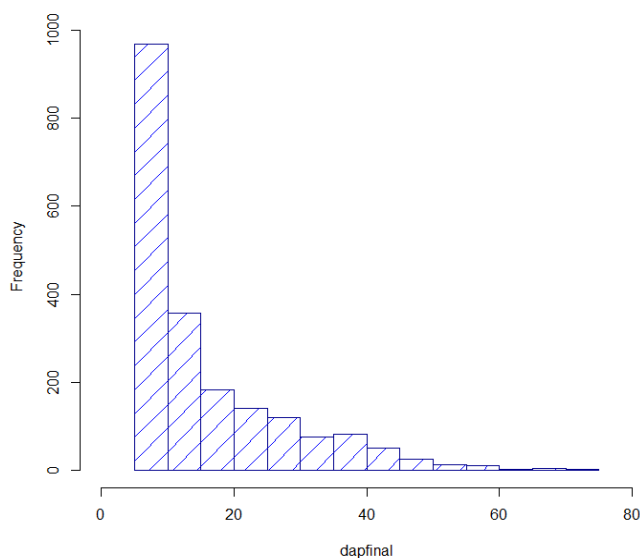
A Tabela 4.1 apresenta os valores mínimo, máximo, médio, quartis e de dispersão (desvio padrão e coeficiente de variação) para as variáveis preditoras.

Observa-se na Tabela 4.1, que a variável Dap_i possui grande variabilidade, caracterizada pelo valor do CV, de 88,26%, e por sua amplitude. É possível observar que o valor da média, primeiro e terceiro quartis do diâmetro são próximos ao seu valor mínimo, o que significa que existem muitas unidades amostrais com diâmetro pequeno. Esse comportamento é esperado em florestas naturais caracterizando uma distribuição diamétrica que se enquadra no modelo exponencial negativo, ou chamado de J-invertido conforme pode ser visualizado na figura 4.1 (DURIGAN, 2009; SOARES; PAULA NETO; SOUZA, 2011).

Tabela 4.1: Estatísticas descritivas das variáveis independentes quantitativas para as 26.313 unidades amostrais.

Variável	Mínimo	Média	Máximo	Desvio	CV	Q1	Q2	Q3
Dap _i	5,00	13,34	149,50	11,77	88,26	6,50	9,10	15,00
G _{·pj}	4,32	7,87	11,93	1,37	17,39	6,87	7,71	8,56
G _{·spj}	0,01	0,33	2,16	0,26	78,66	0,16	0,26	0,41
P _{·pj}	220,00	322,85	508,00	43,22	13,39	292,00	320,00	355,00
P _{·spj}	1,00	13,82	31,00	3,82	27,65	11,00	14,00	16,00
H' _{pmin}	3,37	3,95	4,49	0,23	5,92	3,87	3,98	4,09
H' _{pmax}	3,38	4,15	4,59	0,21	5,01	4,00	4,19	4,28
H' _{pj}	3,38	4,13	4,59	0,21	5,21	3,98	4,15	4,27
H' _{pmin.norm}	0,56	0,69	0,78	0,04	6,32	0,66	0,69	0,71
H' _{pmax.norm}	0,62	0,72	0,80	0,04	5,23	0,69	0,72	0,74
H' _{pj.norm}	0,58	0,72	0,80	0,04	5,40	0,69	0,72	0,74

Em que: CV % = coeficiente de variação; Q1= primeiro quartil; Q2 = segundo quartil; Q3 = terceiro quartil

Figura 4.1 – Histograma do diâmetro a altura do peito (DAP_{ifinal} (cm)).

A variável área basal da parcela (G_{·pj}) possui grande amplitude entre o valor mínimo e máximo, no entanto grande parte das parcelas possuem valores de (G_{·pj}) concentrados em torno da média (observar média, de 7,87, primeiro quartil, 6,87 e terceiro quartil, 8,56).

O comportamento dos dados referentes à área basal da subparcela ($G_{.spj}$) foi semelhante ao da variável ($G_{.pi}$), diferindo em uma menor dispersão absoluta dos valores com 0,26, que pode ser considerado um valor baixo de dispersão. Sua média foi de 0,33, com valor mínimo de 0,01 e máximo de 2,16.

Quanto à quantidade de unidades amostrais na parcela ($P_{.pi}$) tem-se uma dispersão relativa de 13,39% e média de 323 árvores na parcela. Já a quantidade de unidades amostrais na subparcela ($P_{.spj}$) tem-se uma dispersão relativa de 27,65%, com uma diferença de 30 unidades entre o mínimo e o máximo.

A variável indicativa da diversidade de espécies na floresta H' , elucidou a grande variabilidade existente nas subparcelas e por consequência na competição que incide sobre as unidades. O valor mínimo do índice de Shannon foi de 0,56 com o máximo de 0,80 indicando um nível considerável de diversidade. Esse índice é usado em estudos de recursos genéticos, como uma medida de riqueza e uniformidade, cujo baixo H' indica frequência de classes desbalanceadas para uma determinada característica e falta de diversidade genética (PERRY e MCINTOSH, 1991).

Foram estudados modelos com todas as espécies, que perfazem a 26.313 unidades amostrais, os quais não resultaram em bons resultados. Finalmente, foram selecionadas as seis espécies mais balanceadas quanto à dicotomia (morta e viva) resultando em 2.038 unidades amostrais em 60 parcelas permanentes para o ajuste dos modelos com os referidos fatores de competição, o que equivale a 7,75% do total de unidades amostrais.

As seis espécies estudadas foram três espécies comerciais: *Carapa guianensis*, *Guarea carinata*, e *Jacaranda copaia* e três não comerciais: *Cecropia sciadophylla*, *Guatteria poeppigiana* e *Talisia macrophylla*. A Tabela 4.2 apresenta as quantidades de unidades amostrais por espécie e dicotomia.

Tabela 4.2: Espécies selecionadas em função do balanceamento dicotômico.

Código	Nome Botânico	Nome Vulgar	Viva	Morta	% vivas	Total
1295	<i>Carapa guianensis</i>	Andiroba	144	56	72,00	200
1265	<i>Cecropia sciadophylla</i>	Embaúba-torém	221	190	53,77	411
667	<i>Guarea carinata</i>	Andirobarana-vermelha	105	162	39,33	267
2335	<i>Guatteria poeppigiana</i>	Envira-preta	195	154	55,87	349
1999	<i>Jacaranda copaia</i>	Parapará	396	198	66,67	594
904	<i>Talisia macrophylla</i>	Pitomba-folha-grande	131	86	60,37	217
Total			1192	846	-	2038

A Tabela 4.3 apresenta as estatísticas descritivas por código das espécies com as variáveis que foram utilizadas na modelagem.

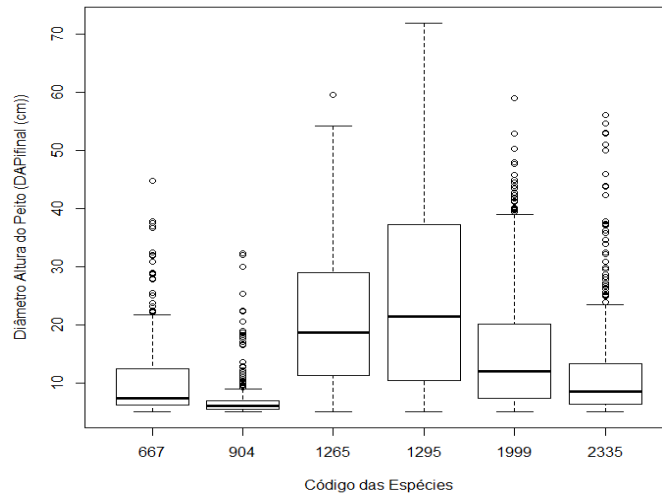
Tabela 4.3: Estatística descritiva das variáveis selecionadas para o modelo logístico final.

Variável	Código	Mínimo	Média	Máximo	Desvio	CV	Q₁	Q₂	Q₃
DAP _{ifinal}	667	5,00	10,66	44,80	7,22	67,75	6,20	7,40	12,50
	904	5,00	7,48	32,30	4,47	59,83	5,40	6,00	6,95
	1265	5,00	21,04	59,60	11,87	56,44	11,30	18,70	29,00
	1295	5,00	25,69	72,00	17,74	69,07	10,40	21,45	37,27
	1999	5,00	15,35	59,00	10,39	67,64	7,40	11,95	20,13
	2335	5,00	12,08	56,10	9,66	79,98	6,30	8,50	13,35
G _{-pfinal}	667	4,98	7,99	11,45	1,27	15,93	6,98	7,74	8,88
	904	4,98	7,70	11,93	1,38	17,93	6,80	7,68	8,45
	1265	4,98	7,96	11,28	1,41	17,76	6,97	7,69	8,88
	1295	4,98	7,74	11,93	1,35	17,47	6,80	7,60	8,55
	1999	4,98	7,85	11,28	1,40	17,82	6,87	7,62	8,58
	2335	4,98	7,54	11,93	1,29	17,15	6,67	7,44	8,20
P _{-spfinal}	667	5,00	13,07	23,00	3,57	27,31	10,00	13,00	15,00
	904	4,00	14,10	25,00	3,95	28,01	11,00	14,00	17,00
	1265	5,00	14,71	28,00	4,39	29,84	11,00	14,00	18,00
	1295	3,00	13,85	26,00	4,09	29,50	11,00	14,00	16,00
	1999	2,00	15,06	28,00	4,21	27,97	12,00	15,00	17,00
	2335	4,00	13,54	28,00	3,89	28,70	11,00	13,00	16,00
H _{-pnom}	667	0,59	0,70	0,80	0,04	5,43	0,68	0,70	0,73
	904	0,60	0,72	0,80	0,04	5,14	0,70	0,72	0,74
	1265	0,58	0,70	0,80	0,04	6,01	0,67	0,71	0,73
	1295	0,62	0,71	0,80	0,03	4,78	0,69	0,72	0,73
	1999	0,59	0,71	0,80	0,04	5,28	0,69	0,71	0,73
	2335	0,59	0,71	0,80	0,04	5,54	0,69	0,71	0,74

Em que: CV % = coeficiente de variação; Q₁ = primeiro quartil; Q₂ = segundo quartil; Q₃ = terceiro quartil.

A Figura 4.2 apresenta o gráfico de boxplot do diâmetro à altura do peito por espécie, onde pode-se observar que as espécies apresentam níveis modais diferentes de diâmetro em virtude das características específicas de cada espécie.

Figura 4.2: Boxplot do diâmetro altura do peito (DAP_{final} (cm)) por espécies.



Todas as espécies apresentaram assimetria positiva o que mostra a juventude das espécies, pois seus diâmetros não atingiram todo o seu ápice de desenvolvimento. A *Guatteria poeppigiana* (2335) apresentou uma maior quantidade de observações aberrantes.

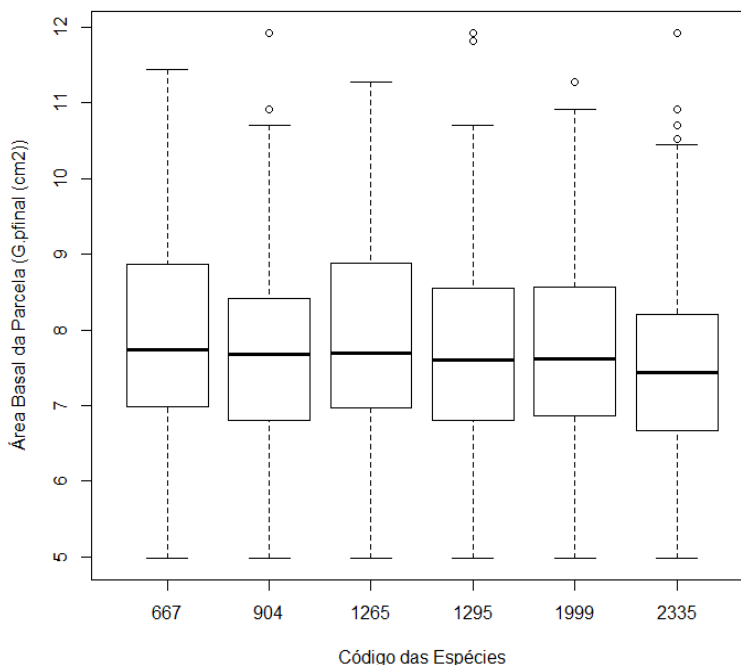
A diferença de comportamento das espécies no ambiente competitivo da floresta reforça a idéia da mortalidade em indivíduos mais jovens, sendo uma característica importante na modelagem, Tabela 4.4

Para esclarecer o resultado apresentado na Tabela 4.3 de que quanto maior o DAP da unidade amostral menor a probabilidade de morrer, a Tabela 4.4 apresenta o *status* das espécies estudadas relativamente às classes de DAP.

Nota-se uma mortalidade muito grande na classe DAP [5 – 10] com 55,7%, o que revela um ponto crítico no desenvolvimento das espécies, sendo a classe de maior probabilidade de morte, ou seja, quanto maior a classe DAP maior a chance das espécies sobreviverem à competição existente na floresta.

Tabela 4.4: Classes comparativas DAP_{ifinal} (cm) para a condição de mortalidade das espécies estudadas.

Estado		Classe DAP_{ifinal} (cm)					Total
		5 -10	10-20	20 - 40	40 - 60	> 60	
Viva	Quantidade	490	316	303	74	9	1192
	%	41,10	26,50	25,40	6,20	0,8	100,00
Morta	Quantidade	471	228	119	27	1	846
	%	55,70	27,00	14,10	3,20	0,1	100,00
Total	Quantidade	961	544	422	101	10	2038
	%	47,20	26,70	20,70	5,00	0,5	100,00

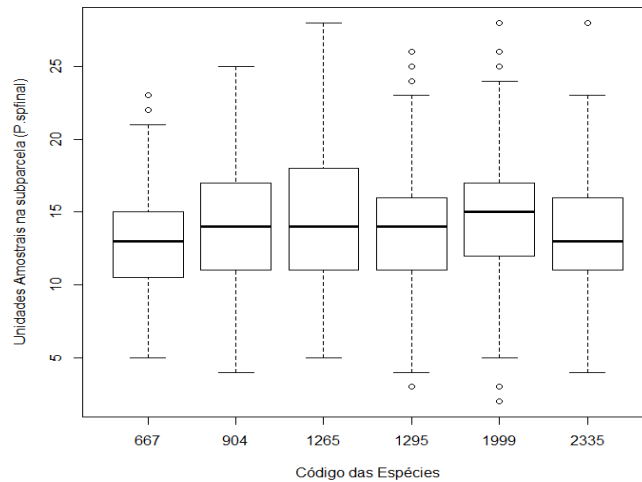
Figura 4.3: Boxplot da área basal da parcela (G_{pfinal} (cm²)) por espécies.

A variável área basal da parcela (G_{pfinal}) apresenta um balanceamento em todas as suas medidas, todas as espécies apresentaram assimetria positiva, poucos pontos são destacados como possíveis *outliers*, sendo novamente a espécie *Guatteria poeppigiana* (2335) a que apresentou uma maior quantidade de observações aberrantes (Figura 4.3).

Como não existe grande variabilidade entre as seis espécies estudadas em relação a área basal da parcela (G_{pfinal}), essa variável podem não ter grande efeito na modelagem.

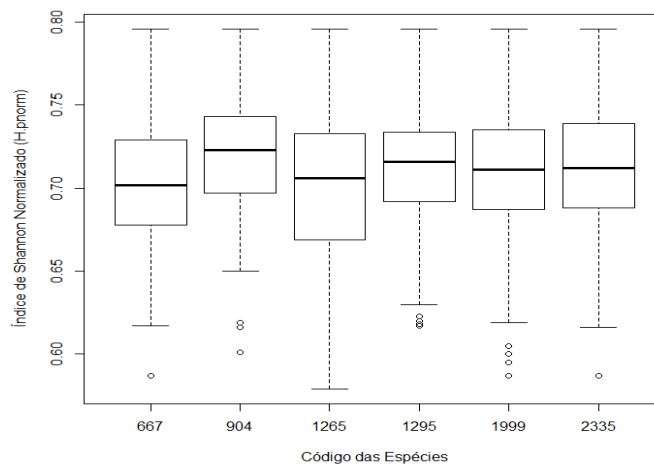
A variável total de árvores na subparcela ($P_{spfinal}$) apresenta uma crescente nas médias que vão da espécie (667) até a espécie (1999). Somente a espécie *Carapa guianensis* (1295) é assimétrica negativa, sendo a *Jacaranda copaia* (1999) aquela que apresentou uma maior quantidade de pontos extremos em seu limite superior (Figura 4.4).

Figura 4.4: Boxplot das unidades amostrais na subparcela (P_{spfinal}) versus código das espécies.



A variável diversidade obtida por meio do índice de Shannon apresenta balanceamento nos resultados, cinco espécies apresentam assimetria negativa, sendo que a espécie (2335) a única com assimetria positiva. A espécie *Jacaranda copaia* (1999) apresentou dezessete pontos extremos em seu limite inferior (LI), pontos esses fora dos chamados valores adjacentes do boxplot (Figura 4.5).

Figura 4.5: Boxplot do índice de Shannon (H_{pnorm}) por espécies.



Os tratamentos silviculturais são as intervenções adotadas na tentativa de redução da concorrência entre as unidades amostrais plantadas, além dessas intervenções existem outras de caráter preventivo, ou seja com o intuito de evitar danos físicos ou fisiológicos as unidades arbóreas. O plano de proteção florestal deve adotar um calendário de vistorias periódicas e medidas a serem tomadas (CALDEIRA, 1999).

Cada floresta requer um sistema silvicultural próprio, não existindo um sistema que possa ser aplicado indistintamente, os diferentes parâmetros existentes para cada floresta podem não ser a ela adequados. O diâmetro mínimo de exploração, o diâmetro de raleamento e outras variáveis não podem ser iguais para florestas com diferentes distribuições diamétricas (SOUZA, JARDIM, 2010).

Na tabela 4.5 e exibida a quantidade de mortalidade existente por tratamento, assim para o tratamento 1 é observado uma mortalidade grande quando comparada com os demais os tratamentos.

Tabela 4.5: Quantidade de unidades amostrais vivas e mortas por tratamento.

Tratamentos	0	1	Total
0 (controle)	119	54	173
1	210	241	451
2	233	153	386
3	248	201	449
4	382	197	579
Total geral	1192	846	2038

4.2.2 Modelagem da mortalidade por meio da regressão binária.

O processo de modelagem inclui o teste de significância dos coeficientes das variáveis independentes, a seleção das covariáveis no modelo e a avaliação do ajuste. Inicialmente é feita a análise de regressão univariada, na qual a significância do coeficiente estimado de cada covariável é verificada pelo teste de Wald.

No banco todas as 12 covariáveis foram testadas, sendo gerados os respectivos coeficientes estimados, Odds Ratio bruto (OR bruto), o intervalo de confiança para o OR (IC-95%), erro padrão e significância no teste de Wald (p-valor).

Nessa etapa da análise, cada covariável foi testada com a variável dependente utilizando o método enter, onde se obtiveram a razão de chances (Odds Ratio), o intervalo de confiança para a razão de chances, o p-valor do teste Wald e o erro padrão estimado de cada covariável. Utilizando o critério sugerido por Hosmer e Lemeshow (1989) a variável que apresentou o valor de $p \leq 0,25$ no teste de Wald, foi considerada candidata a compor o modelo de regressão múltipla.

Concluídas as etapas anteriores, foram utilizados três métodos para a seleção das covariáveis no modelo de regressão múltipla, que foram: forward, backward e stepwise. A significância dos coeficientes das variáveis independentes na regressão múltipla foi verificada pelo valor de p do teste de Wald. Utilizou-se como critério para a permanência da variável independente no modelo final de regressão múltipla p-valor menor que 5% pelo teste de Wald.

No método enter todas as covariáveis foram inseridas, simultaneamente, no modelo logístico e a eliminação de variáveis não significativas foi feita manualmente, ou seja, uma de cada vez e novamente o modelo era ajustado, até o momento que restassem apenas as variáveis significativas pelo teste de Wald no modelo múltiplo.

No método *forward* o modelo inicial foi composto apenas com a constante e a seguir cada passo, foram inseridas uma a uma as variáveis independentes significativas pelo valor de score e removidas baseado na probabilidade do teste de Wald.

Já no método *backward* o modelo inicial foi composto por todas as variáveis independentes e passo a passo, cada variável não significativa no teste de Wald foi removida. O método *stepwise* é automático e realiza o procedimento de modelagem passo a passo.

Dessa forma, foi obtido o modelo de regressão múltipla com as covariáveis selecionadas para a mortalidade das árvores. A significância do modelo testada pelo método de Pearson, Deviance e Hosmer e Lemeshow, assim como a medida de associação (variável resposta e as probabilidades previstas).

A análise revelou as variáveis DAP_{ifinal} , G_{pfinal} , $P_{spfinal}$, H'_{pnorm} , $Trat$, Cod são importantes para o modelo, assim também foi consistente a interação entre as variáveis $P_{spfinal}$ versus H'_{pnorm} . As demais variáveis não obtiveram a significância exigida pelo teste de Wald, portanto seus coeficientes estimados diferem do esperado, e assim foram definidas, as variáveis que serão testadas no modelo de regressão binária múltipla.

4.2.3 Qualidade do ajuste do modelo binomial com suas funções de ligação

O modelo final de regressão binário multivariado foi gerado para cada uma das funções de ligação (logit, probit e cloglog), como não houve grandes mudanças nos parâmetros estimados optou-se por trabalhar com a função de ligação logística.

De posse do modelo final, diversos resultados são gerados pelo programa com *deviance* residual igual a 2355,1 com 2023 graus de liberdade e AIC igual a 2385,1. O valor da estatística G foi 410,376, com valor de p igual a 0,000, indicando que há evidência suficiente que ao menos um dos coeficientes é diferente de zero.

O teste de Pearson e o teste de Hosmer-Lemeshow não foram significativos com os valores de P iguais a 0,410 e 0,091 respectivamente, superando o nível de significância de 5%, isto significa que o modelo da indicações de um bom ajuste.

Pode-se observar que o DAP_{ifinal} apresenta um impacto negativo sobre a chance de morte indicando que quanto maior o DAP_{ifinal} menor a probabilidade de morte da unidade amostral, essa característica será investigada em maiores detalhes. O total de árvores na subparcela ($P_{spfinal}$) tem coeficiente estimado positivo, significando que, quanto maior o número de árvores na subparcela, maior a chance de morte das espécies em virtude da competição ali existente.

Relativamente aos tratamentos verifica-se que os tratamentos 2, 3 e 4 apresentam chance de mortalidade natural reduzida em 14%, 10% e 51%, respectivamente quando comparados com a área de controle (tratamento 0). Isso pode indicar que os tratamentos silviculturais adotados, com a redução da área basal, diminuindo a concorrência por nutrientes e reduzindo a mortalidade natural (OLIVEIRA, 2005).

Para as espécies estudadas, evidencia-se uma redução na chance de mortalidade das espécies *Talisia macrophylla*, *Carapa guianensis*, *Jacaranda copaia* e *Guatteria poeppigiana* em relação à espécie *Guarea carinata*, entretanto não há indicação de diferença estatisticamente significativa entre a espécie *Cecropia sciadophylla* e a espécie *Guarea carinata* quanto à chance de mortalidade.

O efeito significativo da variável que representa a interação entre o total de árvores na subparcela ($P_{spfinal}$) e a diversidade na parcela normalizada (H_{pnorm}) indica que há uma redução na mortalidade dessas espécies de acordo com o aumento da quantidade de árvores na subparcela a cada aumento do índice de diversidade de Shannon. A Tabela 4.6 apresenta as estimativas do modelo com a função de ligação logit.

Tabela 4.6: Modelo ajustado para os dados de mortalidade da regressão logística múltipla, com as estimativas, erro, Z e P, considerando a função de ligação logit.

Preditor	Coefficiente	Erro	z	p	OR	Inferior	Superior
Constante	7,88	3,63	2,17	0,030			
DAP _{ifinal}	-0,03	0,01	-6,51	0,000	0,97	0,96	0,98
G _{pfinal}	0,01	0,04	0,28	0,781	1,01	0,93	1,09
P _{spfinal}	0,62	0,25	2,52	0,012	1,87	1,15	3,03
H _{pnorm}	-9,64	5,07	-1,90	0,057	0,00	0,00	1,34
Tratamento							
1	-0,34	0,23	-1,48	0,138	0,71	0,46	1,11
2	-0,59	0,22	-2,64	0,008	0,55	0,36	0,86
3	-0,54	0,22	-2,45	0,014	0,58	0,38	0,90
4	-1,18	0,23	-5,04	0,000	0,31	0,19	0,49
Código							
904	-0,66	0,20	-3,25	0,001	0,52	0,35	0,77
1265	-0,26	0,19	-1,39	0,163	0,77	0,54	1,11
1295	-0,79	0,22	-3,54	0,000	0,45	0,29	0,70
1999	-0,88	0,17	-5,24	0,000	0,41	0,30	0,58
2335	-0,43	0,18	-2,41	0,016	0,65	0,46	0,92
P _{spfinal} *H _{pnorm}	-0,87	0,35	-2,48	0,013	0,42	0,21	0,83

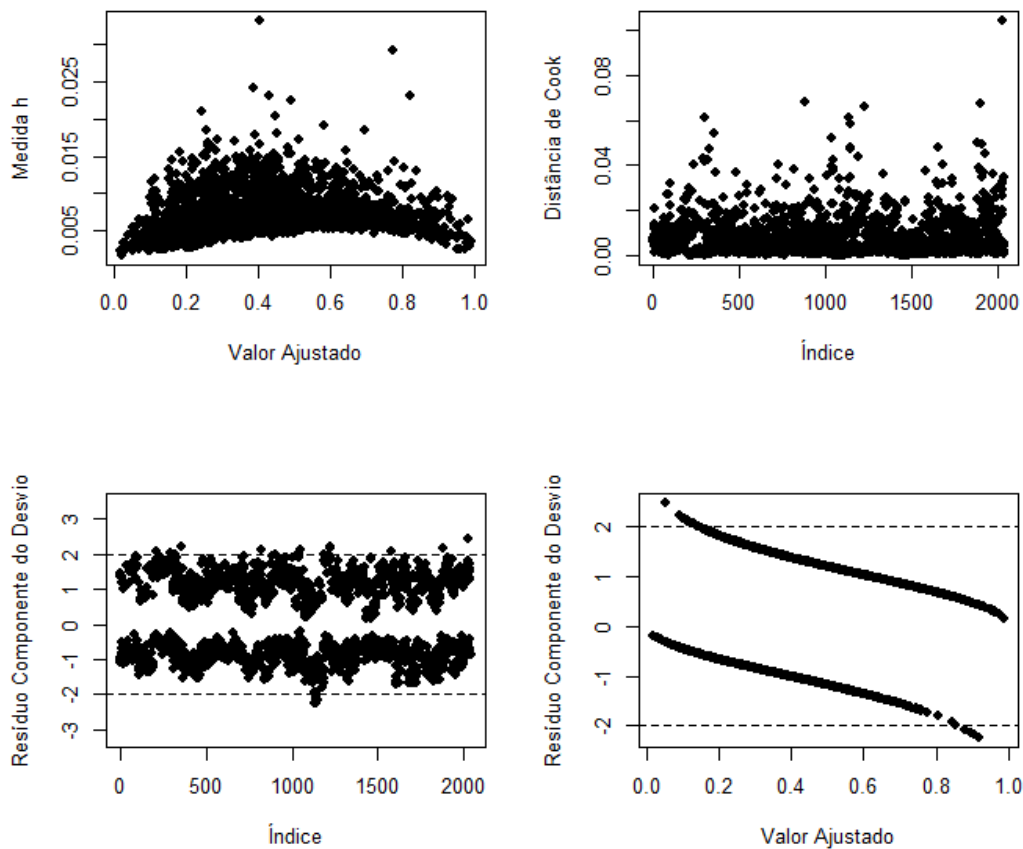
Valores em negrito são significativos considerando $\alpha=5\%$.

A capacidade de prever se uma unidade amostral irá morrer ou não em anos seguintes pode indicar a necessidade de manejos florestais para a sua preservação. A predição da mortalidade de espécies é uma difícil tarefa, já que a morte (natural) é um evento raro (WYCKOFF; CLARK, 2002). A grande diversidade de espécies e diferentes comportamentos dificultam ainda mais a classificação.

Os gráficos apresentados na Figura 4.6 fazem referência a alguns diagnósticos. No canto superior esquerdo pode-se observar o gráfico \hat{h}_{ii} contra valores ajustados onde foi notado um pontos com destaque, (1032). No gráfico de influencia LD_i (distância de cook) destaca-se a observação (2028), já os gráficos dos resíduos t_{D_i} a maioria dos pontos caem dentro do intervalo $[-2,2]$, apenas a observação (2028) se destaca fora do intervalo, as demais estão muito próximas aos limites.

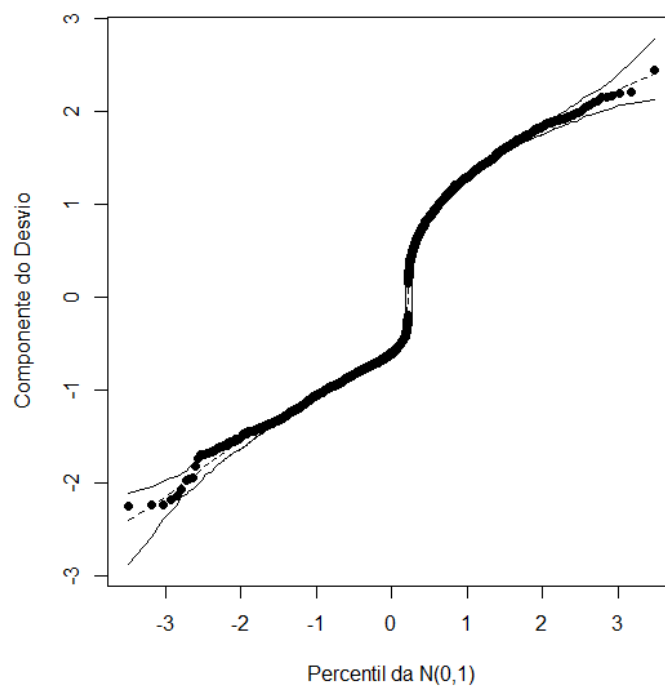
O ponto #1032 corresponde a uma árvore morta *Carapa guianensis*, que ingressou no inventário em 1981 e morreu em 1983, possui um DAP_{final} de 46,00cm, área basal da parcela igual a 7,55cm, P_{spfinal} com 6 unidades amostrais e índice de Shannon normalizado igual a 0,618. O ponto#2028 também corresponde a uma espécie morta *Carapa guianensis* com DAP_{final} igual a 69,50cm tendo seu ingresso em 1983 e morte em 2003.

Figura 4.6: Diagnóstico para um modelo ajustado a dados com com função de ligação logística (logit).



Finalmente a Figura 4.7 apresenta o gráfico normal de probabilidades para o resíduo t_{D_i} , onde não se observa nenhum indício de que a distribuição utilizada seja inadequada. Retirando cada uma das observações destacadas pelos gráficos de diagnostico não foram notadas grandes mudanças inferenciais.

Figura 4.7: Gráfico normal de probabilidade para um modelo ajustado a dados com função de ligação logística (logit).



4.2.4 Treinamento e Validação do Modelo

Para a validação do modelo o banco de dados foi dividido em conjuntos para treinamento e validação, sendo que o conjunto de validação é usado para otimizar os parâmetros. Assim o conjunto de dados foi dividido em subconjuntos de tamanhos proporcionais a quantidade de vivos e mortos bem como a quantidade de cada espécie. Em cada subconjunto foi selecionado aleatoriamente uma parte dos dados para validação e outra para treinamento, na tentativa de reduzir a variância nas estimativas.

4.3 Avaliação da Rede Neural Artificial

Foram feitos arranjos entre os componentes da RNA, optou-se pela seguinte arquitetura: camada de entrada com cinco variáveis, camada intermediária com dez neurônios, e um neurônio na camada de saída (variável dependente - morta ou viva). A função de ativação Tangente Hiperbólica e o algoritmo de treinamento foi o *Back-propagation* de *Levenberg-Marquardt* com taxa de aprendizado 0,005 e o programa computacional utilizado foi o Matlab 6.0 (MATHWORKS, 2009).

A Tabela 4.7 apresenta a taxa global de classificação nos conjuntos de treinamento e validação, bem como a sensibilidade e especificidade nos dois conjuntos de teste, considerando somente as variáveis significativas selecionadas no ajuste da Regressão Logística.

Observa-se que a classificação melhora quando são utilizadas as variáveis significativas pelo modelo de Regressão Logística. Por meio da matriz de confundimento são encontradas a taxa global de classificação nos conjuntos de treinamento e validação, bem como a sensibilidade e especificidade. Considerando a implementação da Rede Neural Artificial se tem um melhor desempenho com o ponto de corte de 0,5.

Foram utilizados três pontos de corte para efeito comparativo (0,4; 0,5 e 0,6) do treinamento e da validação da regressão logística e RNA. Foram feitos alguns testes para melhoria das estimativas e assim obtiveram-se os resultados dados na tabela 4.7.

Tabela 4.7: Taxas de acerto para a regressão logística e RNA das variáveis significativas do modelo, tanto para treinamento quanto para validação.

Técnica	Fase	Taxas			Ponto de Corte
		TG	TS	TE	
Regressão Logística	Treinamento	67,70	69,40	66,40	0,4
	Validação	69,40	64,80	72,30	
	Treinamento	68,40	62,20	80,30	0,5
	Validação	69,80	59,70	82,40	
	Treinamento	67,90	56,80	81,00	0,6
	Validação	70,20	56,60	81,50	
Rede Neural Artificial	Treinamento	70,61	63,90	76,50	0,4
	Validação	71,00	65,80	74,30	
	Treinamento	70,20	70,70	71,50	0,5
	Validação	71,40	65,40	74,30	
	Treinamento	68,90	74,90	67,10	0,6
	Validação	70,60	69,20	71,10	

A RNA teve para este problema um melhor desempenho discriminatório quando comparado com a regressão logística binária. É importante salientar que apesar do melhor desempenho obtido com as redes neurais artificiais, as duas metodologias cumpriram os objetivos a que se propuseram, garantindo acertos entre 67% e 72% na taxa global, indicando um poder classificatório razoável.

Capítulo 5

CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES

Este capítulo apresenta na Seção 5.1 as considerações finais e na 5.2 algumas recomendações para trabalhos futuros.

5.1 Considerações finais

Os MLG ajustaram os dados de mortalidade ao modelo binomial e função de ligação logística, sendo justificável o seu uso para gerar estimativas, principalmente porque não necessitaram de transformação nas variáveis.

A Regressão Logística como modelo de previsão sugere que as variáveis diâmetro altura do peito, número de unidades amostrais na subparcela, tratamento, as espécies e a interação entre número de unidades amostrais e diversidade são importantes para o estudo da mortalidade das espécies arbóreas estudadas.

Essa técnica permitiu definir as variáveis relevantes do problema e que foram utilizadas na Rede Neural Artificial. A Regressão Logística pode ser utilizada como uma primeira etapa na análise, com o objetivo de selecionar os fatores associados à mortalidade, subsidiando a implementação da solução.

O desempenho obtido pela Rede Neural Artificial foi de grande relevância, pois se obteve uma melhora na classificação dos dados a partir das variáveis de competição especificadas no presente estudo.

5.2 Recomendações

Recomenda-se para trabalhos futuros:

- i)** Aplicar a metodologia utilizada nesta dissertação para estudar outros fatores que possam englobar mais covariáveis, como por exemplo (variáveis climáticas, variáveis de solo).
- ii)** Utilizar a geoestatística para georreferenciar a floresta, de maneira a monitorá-la de forma mais eficaz.
- iii)** Tentar agrupar espécies de forma a gerar menor competição entre elas, diminuindo assim a mortalidade por competição, seja ela por nutrientes, luz e espaço.
- iv)** Aplicar outras funções de ligação e compará-las buscando um melhor desempenho. Aumentando o tamanho da amostra para uma melhora da qualidade do ajuste.

REFERÊNCIAS BIBLIOGRÁFICAS

AKAIKE, H. A new look at statistical model identification. *IEEE Transactions on Automatic Control* **AU-19** 716-722, 1974.

AYALEW, L., YAMAGISHI, H. The application of gisbased logistic regression for landslide susceptibility mapping in the kakuda-yahikomountains, central japan. *Geomorphology*, 65, 15–31, 2005.

BERG, C.C.; ROSSELLI, P.F. *Cecropia*. New York: Flora Neotropica / The New York Botanical Garden, 2005.

CALDEIRA, S. F. PRÁTICAS SILVICULTURAIS, notas de aulas teóricas. Universidade Federal do Mato Grosso, 1999.

CANTERLE, D. R., BAYER, F. M. Test for link misspecification in generalized linear models for binary data, *Ciência e Natura*, Santa Maria, v. 37 n. 1, 2015, p. 1–11 *Revista do Centro de Ciências Naturais e Exatas - UFSM*

CARVALHO, J.O.P. de. Estrutura de matas altas sem babaçu na Floresta Nacional do Tapajós. In: SILVA, J. N.M.; CARVALHO, J.O.P. de; YARED, J.A.G. (Ed.) **A silvicultura na Amazônia Oriental**: contribuições do projeto Embrapa/DFID. Belém: Embrapa Amazônia Oriental/DFID, p.277-290, 2001.

CARVALHO, J.O.P. de. Structure and dynamics of a logged over Brazilian Amazonian rain forest. Oxford, 215 p. Thesis (Ph.D.) - University of Oxford., 1992.

CAVALCANTE, Y, L. Aplicação das Técnicas Análise Multivariada e Redes Neurais Artificiais na Classificação das Águas de Reservatórios de Hidrelétricas: um Estudo de Caso na Região Amazônica. Dissertação de Mestrado em Ciência da Computação. Universidade Federal do Pará, 2013.

CHAVES, F. J. M. Um Método para Previsão de Óbitos de Bebês Utilizando Regressão Logística e Redes Neurais Artificiais: Um Estudo de Caso. Dissertação de Mestrado em Estatística. Universidade Federal do Pará, 2009.

CORDEIRO, G. M.; DEMÉTRIO C. G. B. Modelos Lineares Generalizados e Extensões. Departamento de Estatística e Informática – UFRPE/Departamento de Ciências Exatas – USP, 2013.

DIAS FILHO J. M; CORRAR, L. J. Regressão Logística.p.281-323 In: CORRAR; L. J. etal. (coord.) Análise Multivariada: para os cursos de administração, ciências contábeis e economia. Atlas, São Paulo, 2007.

DURIGAN, G. Estrutura e Diversidade de Comunidades Florestais. In: MARTINS,S.V. (Ed.) Ecologia de florestas tropicais do Brasil. Editora UFV. UniversidadeFederal de Viçosa. 2009. 261 p

EID, T., TUHUS, E., Models for individual tree mortality in Norway. Forest Ecology and Management 154:69-84. 2001.

EMBRAPA. Espécies Arbóreas da Amazônia. Nº 02: AndirobaCarapaguianensis. s/p 2004. Disponível em: [HTTP://www.cpatu.embrapa.com.br/dendro,tudo1.htm](http://www.cpatu.embrapa.com.br/dendro,tudo1.htm)

Fonte das informações nomenclaturais e de distribuição por UF. Flora do Brasil. Jardim Botânico do Rio de Janeiro. Flora do Brasil Disponível em: <http://www.floradobrasil.jbrj.gov.br>

Fonte de informações <https://portaldabiodiversidade.icmbio.gov.br> Instituto Chico Mendes, ICMBio, 2017. Acessado em Maio de 2017.

Fontedeinformações:<https://portaldabiodiversidade.icmbio.gov.br/portal/occurrences/c6d44308-04b0-4b99-a6de-2bcce4779e8b>. Acessado em 05/04/2017.

GAGLIOTI, A. L. Urticaceaejuss no Estado de São Paulo, Brasil. Dissertação (Mestrado em Biodiversidade Vegetal e Meio Ambiente) - Instituto de Botânica, Secretaria de Estado do Meio Ambiente, São Paulo, 195 p, 2011.

GONÇALVES, V. A. Levantamento de mercado de produtos florestais não madeireiros. Santarém IBAMA – Promanejo, 2001.

HOSMER, D. W. Jr.; LEMESHOW, S. Applied logistic regression. New York: Wiley, 1989.

KING, S.; BENNETT, K.; LIST, S. Modeling noncatastrophic individual tree mortality using logistic regression, neural networks, and support vector methods.Computers and Electronics in Agriculture, v. 27, n. 1-3, p.401-406, 2000.

LORENZI, H. Árvores brasileiras. Manual de identificação e cultivo de plantas arbóreas nativas do Brasil. Nova odessa. Ed. Plantarum 1992.

LORENZI, H. Árvores brasileiras: manual de identificação e cultivo de plantas arbóreas do Brasil 2. ed. Nova Odessa: Instituto Plantarum, 381 p. Vol. 2. , 2002.

- MARCIANO F. W. P. Principais tipos de resíduos utilizados na análise de diagnóstico em MLG com aplicações para os modelos: Poisson, ZIP e ZINB. Projeto de Iniciação Científica. Departamento de Estatística e Matemática Aplicada – UFCE. 2009.
- MAUÉS, M. M. Estratégias reprodutivas de espécies arbóreas e sua importância para o manejo e conservação florestal: Floresta Nacional do Tapajós (Belterra-PA). 2006. Tese de doutorado em ecologia – Universidade Federal de Brasília, 2006.
- McCULLAGH, P., NELDER, J. Generalized linear models, 2^o edn. Chapman and Hall. 1989.
- MONSERUD, R.A.; STERBA, H. Modeling individual tree mortality for Austrian forest species. *Forest Ecology and Management*, v.113, p.109-123, 1999.
- NEGNEVITSKY, M. Artificial Intelligence: A guide to intelligent systems. 2. ed.[S.l.]: Pearson Education, 415 p.,2005.
- OLIVEIRA S. Inferência e análise de resíduos e de diagnóstico em modelos lineares generalizados. Monografia. Graduação em Estatística – UFJF,2013.
- OLIVEIRA, L. C. de; COUTO, H. T. Z. do; SILVA, J. N. M.; CARVALHO, J. O. P. de. Efeito da exploração de madeira e tratamentos silviculturais na composição florística e diversidade de espécies em uma área de 136ha na Floresta Nacional do Tapajós. *Scientia Forestalis*, n. 69, p. 62-76, 2005.
- ÖZÇELİK, R.; DIAMANTOPOULOU, M. J.; CAMPO, F. C.; ELER, ÜNAL. Estimating Crimean juniper tree height using nonlinear regression and artificial neural networks models. *Forest Ecology and Management*, Vol. 306, p. 52-60, October 2013.
- PALMA NETO, L. G; NICOLETTI, M. C. Introdução as Redes Neurais construtivas. São Carlos/SP: Editora da Universidade Federal de São Carlos, 2005.
- PAULA G. A. Modelos de regressão com apoio computacional. Instituto de Matemática e Estatística – USP,2010.
- PERRY, M. C.; MCINTOSH, M. S. Geographical patterns of variation in the USA soybean germplasm collection: I. Morphological traits. *Crop Science*, Madison, v. 31, n. 5, p. 1350- 1355, Sept./Oct. 1991.
- Pereira Júnior, O.L.; Wolter Filho, W. ; da Rocha, A F. .1.; Carvalho, M.G. ; Braz Filho, R., 1990. Diterpenos isolados de *Guarea carinata*. *Química Nova* 13, pp. 247-249.

PHILLIPS, D. P.; AZEVEDO, C. P. de; DEGEN, B.; THOMPSON, I. S.; SILVA, J.N.M.; GARDINGEN, P.R. van. An individual-based spatially explicit simulation model for strategic forest management planning in the eastern Amazon. *Ecological Modelling*. 173, 335-354, September, 2003.

PIO-CORRÊA, M. Dicionário das Plantas Úteis do Brasil e das Exóticas Cultivadas, Rio de Janeiro: Imprensa Nacional, p. 200-212, 1978.

QUEIROZ, W. T. Amostragem em inventário florestal. Universidade Federal Rural da Amazônia, 2012.

RAMSEY, D. S. L.; FORSYTH, D. M.; VELTMAN, C. J.; NICOL, S. J.; TODD, C. R.; ALLEN, R. B.; ALLEN, W. J.; BELLINGHAM, P. J.; RICHARDSON, S. J.; JACOBSON, C. L.; BARKER, R. J. An approximate Bayesian algorithm for training fuzzy cognitive map models of forest responses to deer control in a New Zealand adaptive management experiment. *Ecological Modelling*, Vol. 240, p. 93-104, 2012.

RIBEIRO, J. E. L. S. et al. 1999. Flora da reserva Ducke: Guia de identificação de plantas vasculares de uma floresta de terra firme na Amazônia Central. Manaus: INPA.

RODRIGUES, A. P. Melicocceae (Sapindaceae): *Melicoccus* and *Talisia*. *Flora Neotropica* 87:1-178, 2003.

ROSSI, L.M.B.; et al. Modelagem da mortalidade em florestas naturais. *Revista Floresta*, v.37, p.275-291, 2007.

SFB – SERVIÇO FLORESTAL BRASILEIRO. Florestas do Brasil em resumo – 2010: dados de 2005 – 2010. Brasília: SFB, 2010.

SILVA, R.C.V.M. Conhecendo Espécies de Plantas da Amazônia (*Cecropiasciadophylla* - *Urticaceae*). Comunicado técnico, 234, 2012.

SOARES, C. P. B.; PAULA NETO, F.; SOUZA, A. L. Dendrometria e inventário florestal. Viçosa, MG: Universidade Federal de Viçosa, 2 ed. 272p. 2011.

SOUZA, A. L. ; JARDIM, F. C. S. SISTEMAS SILVICULTURAIS APLICADOS ÀS FLORESTAS TROPICAIS, Manejo Florestal – DEF/UFV, 2010.

The R foundation for statistical computing; R Program; A Free Software for Statistical Computing; Vienna, Austria, 2014.

VANCLAY, J. K. Mortality functions for north Queensland rainforests. *Journal of Tropical Forest Science* 4:15-36, 1991,

WEDDERBURN, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika* 68, 27-32.

WYCKOFF, P. H.; CLARK, J. S. The relationship between growth and mortality for seven co-occurring tree species in the southern Appalachian Mountains. *Journal of Ecology*, Oxford, v. 90, n. 4, p. 604-615, 2002.

YANG, Y.; TITUS, S. J.; HUANG, S. Modeling individual tree mortality for white spruce in Alberta. *Ecological Modeling*, Amsterdam, v. 163, n. 3, p. 209-222, 2003.

YAO, X., TITUS, S.J., MACDONALD, S.E. A generalized logistic model of individual tree mortality for aspen, white spruce, and lodgepole pine in Alberta mixedwood forests. *Canadian Journal of Forest Research*, 31 (2), 283-291.2001.

MATHWORKS, Neural Network Toolbox User's Guide: for use MATLAB, Matlab6p1, 2009.