UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
INSTITUTO DE MATEMÁTICA
PROGRAMA DE PÓS GRADUAÇÃO
MESTRADO PROFISSIONAL
MATEMÁTICA EM REDE NACIONAL

THIAGO FARIAS MACÊDO ARCE

REGRESSÃO LOGÍSTICA APLICADA A PECUÁRIA DE CORTE

CAMPO GRANDE - MS
2017

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
INSTITUTO DE MATEMÁTICA
PROGRAMA DE PÓS GRADUAÇÃO
MESTRADO PROFISSIONAL
MATEMÁTICA EM REDE NACIONAL

THIAGO FARIAS MACÊDO ARCE

REGRESSÃO LOGÍSTICA APLICADA A PECUÁRIA DE CORTE

Orientadora: Profª. Drª. Rúbia Mara de Oliveira Santos

Dissertação apresentada ao Programa de Pós-Graduação Mestrado Profissional em Matemática em Rede Nacional do Instituto de Matemática da Universidade Federal de Mato Grosso do Sul-INMA/UFMS como parte dos requisitos para obtenção do título de Mestre.

CAMPO GRANDE - MS
2017

REGRESSÃO LOGÍSTICA APLICADA A PECUÁRIA DE CORTE

THIAGO FARIAS MACÊDO ARCE

Dissertação apresentada ao Programa de Pós-Graduação Mestrado Profissional em Matemática em Rede Nacional, do Instituto de Matemática da Universidade Federal de Mato Grosso do Sul-INMA/UFMS como parte dos requisitos para obtenção do título de Mestre.

Aprovado pela Banca Examinadora:

Profª. Dra. Rúbia Mara de Oliveira Santos -UFMS
Prof. Dr. Roberto Quirino do Nascimento - UFPB
Prof. Dr. Erlandson Ferreira Saraiva - UFMS

Campo Grande – MS, Julho de 2017

All models are wrong, but some are useful.

George Box

Agradecimentos

Cada vez que me aprofundo no estudo da matemática, em especial nessa etapa estudando estatística, é natural me sentir grato ao Criador Jeová pela forma como o universo é organizado e como Ele nos dá a inteligência para seguir aprendendo cada vez mais de sua criação.

Agradeço a minha familia, minha esposa Erbelle, fiel companheira de minhas manhãs, tardes e noites dedicadas na pesquisa e escrita desse trabalho. Foi e sempre será minha fonte de motivação para cada luta, me garantindo sua ajuda a ter uma visão equilibrada das coisas.

Esse agradecimento se estende a toda minha familia. Aos meus pais Reinaldo e Tânia e irmã Letícia pelo lar que tive onde pude desenvolver meu melhor. Aos meus sogros Macêdo e Erbene que nunca duvidaram de meu potencial, mesmo quando eu mesmo fui incrédulo.

À professora Dra. Rúbia Mara de Oliveira Santos estendo meus agradecimentos por sua experiência, conselhos e tempo dedicados na contribuição desse texto, sempre visando a excelência, me incentivando sempre a melhorar.

Ao professor Dr. Erlandson Ferreira Saraiva agradeço por suas aulas onde pude ter o primeiro contato real com a probabilidade e estatística. Além disso suas inúmeras sugestões demonstraram ser de imensa valia.

Agradeço a professora Dra. Elisabete Souza Freitas por suas aulas durante toda minha formação como matemático. A considero uma modelo que levo na minha prática docente.

Agradeço a colaboração do LSCAD, o professor Dr. Ricardo e ao Leonardo por sua ajuda e disposição, que tornaram possível a viabilização desse trabalho com sua ajuda técnica.

Aos professores do INMA pelos conhecimentos transmitidos.

Resumo

O estresse térmico de bovinos de corte é um assunto de interesse para pesquisadores e investidores da área. Assim, foi desenvolvido um sistema para captar e indicar o nível de sombreamento em que o animal se encontra, e como consequência, o estado de estresse térmico do animal. Nesse trabalho, uma extensão da regressão logística, a regressão logística multinomial, é utilizada para a obtenção de modelos que indiquem do nível de sombreamento em que bovinos de corte se encontram. Utilizando um procedimento de seleção de modelos, o AIC e o BIC, foram identificadas variáveis que melhor expliquem o nível de sombreamento. Obtendo, como resultado, um modelo com alto poder de previsão e com menor quantidade de parâmetros que um modelo saturado.

Palavras-chave: Regressão Logística Multinomial, Modelos Lineares Generalizados, Estresse Térmico.

Abstract

The thermal stress of beef cattle is a subject of interest to researchers and investors in the area. Thus, a system was developed to capture and indicate the level of shading in which the animal is, and as a consequence, the state of thermal stress of the animal. In this work, an extension of the logistic regression, the multinomial logistic regression, is used to obtain models that indicate the level of shading in which beef cattle are found. Using a model selection procedure, AIC and BIC, variables were identified that best explain the level of shading. Obtaining, as a result, a model with high predictive power and with fewer parameters than a saturated model.

Keywords: Multinomial Logistic Regression, Generalized Linear Models, Thermal Stress.

Nomenclatura e Notações

$\hat{\theta}$	Estimativa de máxima verossimilhança de θ
$\hat{\boldsymbol{\theta}}$	Vetor de estimativas de máxima verossimilhança de $\boldsymbol{\theta}$
$\boldsymbol{\theta}$	Vetor de parâmetros de uma distribuição de probabilidades
$\mathbb{P}(.)$	p-valor
$\mathcal{F}$	Classe de Eventos
$\mathcal{P}(\Omega)$	Partes de $\boldsymbol{\Omega}$
$\mathcal{X}^2_n$	Distribuição qui-quadrado com n graus de liberdade
μ	Valor esperado ou esperança
Ω	Espaço Amostral
ω	Ponto Amostral
π	Medida de probabilidade
σ	Desvio-padrão
σ^2	Variância
Θ	Espaço paramétrico
θ	Parâmetro de uma distribuição de probabilidades
ε	Desvio em relação à média
$A \cap B$	Interseção de um evento A com B
$A \cup B$	União de um evento A com B
A^c	Complementar de um evento A
$E(X)$	Valor esperado ou esperança de uma variável aleatória X
$L(\theta; x)$	Função de verossimilhança
$N(\mu, \sigma^2)$	Distribuição normal de parâmetros μ e σ^2
$(\Omega, \mathcal{F}, P)$	Espaço de Probabilidade
X	Variável preditora independente
Y	Variável resposta dependente

Lista de Tabelas

Lista de Figuras

Sumário

Capítulo 1

Introdução

A etimologia da palavra estatística se origina do vocábulo *status* (estado em latim). A criação da palavra é atribuída ao professor alemão Gottfried Achanwall da Universidade de Göttigen, que deu uma melhor sistematização e definição desse estudo na época da renascença. O primeiro estudo nessa área é datado do século XVII. Desenvolvido na Inglaterra, era denominado como *Aritmética Política* originando o que é hoje conhecido como demografia [1]. Nessa época, o astrônomo Edmond Halley construiu tábuas de ampla utilidade para o cálculo do seguro social elaborada a partir dos registros de vitais da cidade alemã de Bresláu.

Em 1835 Adolphe Quételet, foi pioneiro no uso de métodos quantitativos em sociologia [2]. Conhecido como "pai das estatísticas públicas", Quételet coletou e analisou uma grande quantidade de dados como criminalidade, taxa de divórcios, suicídios, nascimentos, mortes, altura humana, peso, que até então, não se esperava qualquer relação matemática. O estudo de Quételet permitiu chegar a conclusão de que uma massa de pessoas é mais previsível que pessoas individualmente. Ao registrar a proporção de pessoas com determinada altura, obteve uma curva aproximadamente normal. O mesmo acontecia com outras variáveis sociais. A partir de então ele percebeu que a curva de Gauss podia ser ajustada às medidas de peso, estatura e perímetro toráxico mostrando sua aplicabilidade além da distribuição dos erros [2].

O estudo da regressão e correlação começou pelas contribuições de Sir Francis Galton. O termo regressão, foi usado pela primeira vez em 1885 em um comparativo entre pais e filhos. Galton foi um dos fundadores na Inglaterra da conhecida escola Biométrica, que se dedicava a aplicações da estatística à herança biológica [1].

O período que compreende 1890 e 1920 foi um dos mais importantes períodos da história da Estatística com aprimoramento das técnicas de correlação e ajustamento de curvas. Grande

parte das contribuições nesse período se deve a Karl Person considerado o fundador da Estatística Moderna, se dedicando assim como Galton a problemas de estastística relacionados com a herança biológica. No artigo *Regression, Heredity and Panmixia*, Pearson propôs a fórmula para o coeficiente de correlação utilizada até hoje. Seus estudos levaram ao desenvolvimento da regressão e correlação múltiplas e foram base para artigos e trabalhos nessa área. Suas contribuições se extenderam à inferência estatística. Uma dessas contribuições foi o método dos momentos para estimação de parâmetros, e o teste de significância usando a distribuição $\mathcal{X}^2$ (qui-quadrado) para ajustar curvas de frequência.

Comtemporâneo das pesquisas de Person, Ronald Aylmer Fisher contribuiu com o uso do método de máxima verossimilhança para ajustar curvas de frequência. Em um clássico trabalho de 1922, Fisher resolve o problema da estimação pontual usando esse método. O método de máxima verossimilhança é uma das maiores contribuição de Fisher à estatística, permitindo a estimação de parâmetros em modelos tais como a de regressão logística.

O modelo de regressão logística, em especial seu caso multinomial, é foco de estudo no presente trabalho. A função logística foi inventada no século 19 para descrever o crescimento populacional [3]. Para países jovens, estudos descreviam o crescimento populacional como uma curva exponencial, seguindo uma progressão geométrica. Sabendo que um crescimento exponencial levaria a valores impossíveis para uma população, Quételet confiou a Pierre-François Verhulst a tarefa de descrever o crescimento populacional com a inclusão de uma função que descrevesse obstáculos ao crescimento, uma resistência. A adição de uma função de resistência à curva exponencial foi nomeada **função logística** por Verhulst em 1838, sendo redescoberta em 1920 por Raymond Pearl e Lowell Reed.

Em 1944 Joseph Berkson usou a função logística como alternativa ao probit, baseada na curva normal, pelo uso do término **logit** [3]. A função logit demonstrou uma vantagem computacional em relação ao probit, sendo adotada em ciências biológicas, econômicas, de epidemiologia e sociais. Sua generalização multinomial foi primeiramente promovida por Cox em 1966 abrindo potencial para modelar problemas das áreas de economia e de ciências sociais. McFadden, em 1973, ligou o logit multinomial com a teoria da escolha discreta, onde duas ou mais escolhas discretas são possibilidades de escolha. O trabalho de McFadden construiu a base teórica para pesquisas que fazem uso desse modelo de regressão, lhe garantindo um prêmio Nobel no ano de 2000 [3].

Nessa dissertação, modelos de regressão logística multinomial serão propostos com o objetivo

de indicar o conforto térmico de bovinos de corte a partir de variáveis ambientais. Os parâmetros desses modelos serão estimados pelo método de máxima verossimilhança. Dentre os quinze modelos propostos, um deles se destacará devido a seu poder de predição e por obter o melhor resultado em procedimentos adotados para a seleção de modelos.

1.1 O Ensino da Estatística

A Lei de Diretrizes e Bases da Educação Nacional (LDB 9.394/96) [6] divide o ensino no Brasil em duas etapas: a educação básica e o ensino superior, onde a educação básica é subdivida em educação infantil, ensino fundamental e ensino médio.

Segundo a LDB, a educação básica tem como finalidade "desenvolver o educando, assegurar-lhe a formação comum indispensável para o exercício da cidadania e fornecer-lhes meios para progredir no trabalho e nos estudos posteriores"[6].

Para atender esse objetivo os Parâmetros Curriculares Nacionais (PCNs) [4], em conformidade com [6] organiza os conteúdos de matemática da educação básica em 4 blocos: Números e operações; Espaço e forma; Grandezas e Medidas e Tratamento da informação. Tanto [6] como [4] defendem a interdisciplinariedade dos conteúdos contidos nesses blocos.

A estatística está presente no bloco Tratamento da informação. Os Paramêtros Curriculares Nacionais para o Ensino Médio (PCNEM) defendem sua relevância pela sua presença nas ciências exatas, da natureza e humanas [5]. Os PCNs para o ensino fundamental incluem a estatística como de suma importância para a compreensão e a tomada de decisões diante de questões políticas e sociais que envolvem a leitura e interpretação de informações complexas, muitas vezes contraditórias, que incluem dados estatísticos e índices divulgados pelos meios de comunicação [4].

Dos cinco temas transversais relacionados à matemática nos PCNs, a estatística aparece em pelo menos dois deles, a saber, o meio ambiente e a saúde. Em relação ao meio ambiente a estatística é apontada de importância para a compreensão de fenômenos no ambiente. Possibilita que o aluno possua todas as ferramentas necessárias para organizar e interpretar dados, testar e formular hipóteses. Na saúde, esse mesmo documento descreve a importância da estatística para a análise de dados, comportamentos e previsões dessa área. Esse conhecimento permite ao aluno conhecer a si mesmo e compreender aspectos sociais que levam a doenças [4].

O Referencial Curricular da Rede Estadual de Ensino de Mato Grosso do Sul e o Referencial

Curricular da Rede Municipal de Ensino de Campo Grande (REME) são documentos que contêm o currículo de matemática da Rede Estadual de Ensino de Mato Grosso do Sul e da Rede Municipal de Ensino de Campo Grande/MS. Os conteúdos que compreendem a área da estatística estão dentro do eixo tratamento da informação. As tabelas (1.1) e (1.2) descrevem com base nos referenciais curriculares a abordagem da estatística nessas redes de ensino. Mais informações podem ser encontradas com detalhes em [7], [8] e [9].

Tabela 1.1: Estatística no Referencial Curricular da REME

Série	Leitura e Interpretação de dados	Construção de Gráficos	Medidas de Tendência Central	Conceitos básicos: Estatística descritiva	Conceitos básicos: Probabilidade
6º Ano	√	√			
7º Ano	√	√			
8º Ano	√	√	√		
9º Ano	√	√	√	√	√

Tabela 1.2: Estatística no Referencial Curricular da Rede Estadual de Ensino do MS

Série	Leitura e Interpretação de dados	Construção de Gráficos	Medidas de Tendência Central	Estatística Descritiva	Probabilidade	Medidas de Variabilidade
6º Ano	√	√				
7º Ano	√	√	√		√	
8º Ano		√	√	√	√	
9º Ano	√	√	√		√	
3º Ano			√	√	√	√

Ambos referenciais curriculares apresentam enfoque do ensino da estatística descritiva, e pouca introdução de ferramentas para a predição de resultados com base em dados coletados, um dos motivos pelos quais os PCN descrevem com tanta importância o estudo da estatística.

As Orientações Curriculares para o Ensino Médio [10] recomendam o ensino da estatística em todos os níveis da educação básica, em especial para o ensino médio. Além da recomendação da ênfase em construção e representação de tabelas e gráficos, é relatada a importância do estudo culminar em uma apresentação de resultados que se apóiam em inferências tomadas da amostra de uma população.

De acordo com [10] o ensino da estatística deve estar associado com o uso de recursos tecnológicos. Os PCNs também ressaltam a importância desses recursos afirmando que seu uso permite criar ambientes onde novas formas de pensar e aprender são possíveis.

Tendo em vista a facilidade dos alunos em relação à tecnologia e o fato de que em toda escola pública existe ao menos uma **sala de recursos tecnológicos**, uma das tecnologias disponíveis para o ensino da estatística são softwares estatísticos. Tais softwares permitem a manipulação de dados, cálculo e visualização gráfica. Cada software apresenta sua própria linguagem e ambiente de programação.

1.1.1 O software estatístico R

Um software de domínio público cujo download é gratuito é o **software estatístico R** [12]. Sendo um software livre disponível na forma de código fonte, é garantido os direitos de uso, modificação e redistribuição do programa, garantindo para todos esses mesmos direitos. Os códigos-fonte do R estão atualmente disponibilizados e gerenciados por um grupo denominado Core Development Team (http://www.rproject.org/contributors.html) [11].

O R é uma linguagem e ambiente para a computação estatística e plotagem de gráficos, que permite a manipulação de dados, cálculo e visualização gráfica. O uso do programa é feito pela entrada de comandos simples que podem ser agrupados para a criação de funções mais complexas. Sua linguagem clara permite usuário, no caso o aluno, tenha entendimento de cada comando executado, tornando todo o processo altamente didático.

O R fornece uma ampla variedade de técnicas estatísticas. Assim enriqueceria aulas de estatística envolvendo a análise de dados bem como sua visualização gráfica. A figura (1.1) demonstra a visualização de gráficos. O uso desse software em sala de aula traz benefícios como a construção de conceitos estatísticos abordados no ensino fundamental e médio através de uma linguagem de programação.

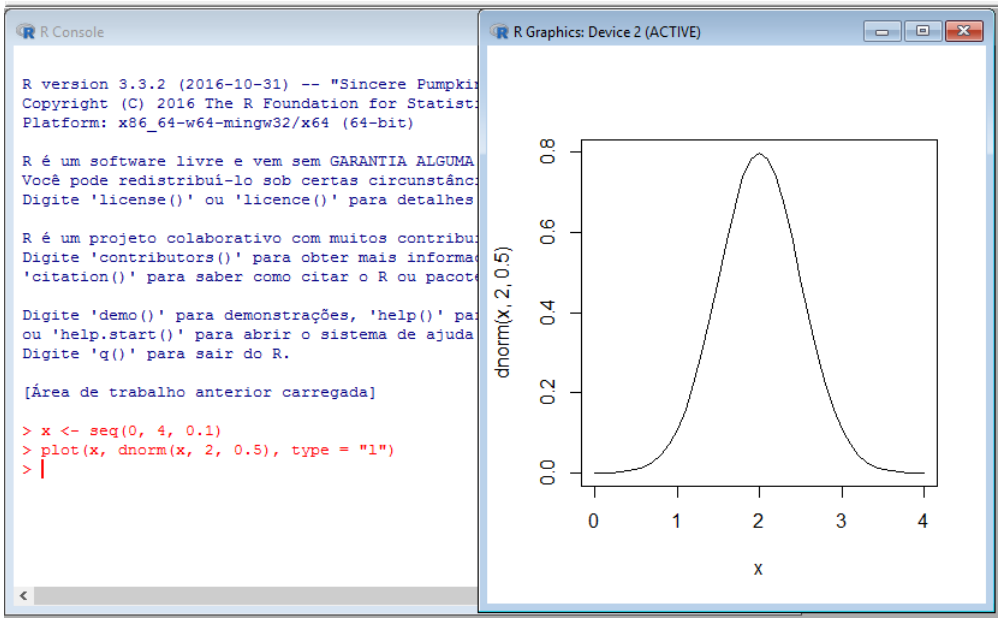

Figura 1.1: Interface e uso do R para visualização

O software estatístico R permite criar ambientes virtuais tornando possível experimentos que seriam impossíveis de serem feitos de modo prático em sala de aula. A figura (1.2) mostra o uso das funções *sample* e *table* para efetuar o lançamento de uma moeda 10^8 vezes, facilitando ensino do conceito de probabilidade por frequência.

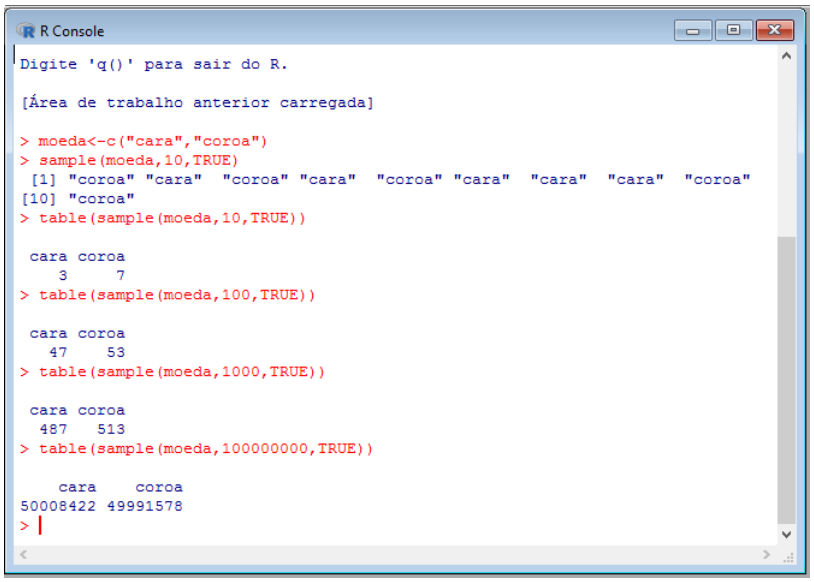

Figura 1.2: Usando o R para grandes amostras

O R é disponibilizado para plataformas como o Unix, Linux, Windows e Mac. Isso permite que a **lousa digital** ou **lousa interativa** possa servir de plataforma para o uso do software

em sala de aula. A lousa digital permite o uso do R no tamanho de uma lousa tradicional possibilitando ao professor a facilidade de executar todos seus comandos com apenas um toque.

Tanto a sala de recursos tecnológicos como a lousa digital munidos do software R, permitem uma aula de estatística ser mais dinâmica, enriquecendo ainda mais seu conteúdo e facilitam ao aluno a construção de conceitos. Algumas limitações do mundo físico, como por exemplo, lançar uma moeda à uma grande quantidade de vezes, podem ser superadas criando novas situações pedagógicas.

1.2 Objetivos

Essa dissertação tem como objetivo propor o modelo de regressão logística multinomial para indicar o nível de sombreamento em que bovinos de corte se encontram a partir de um conjunto de variáveis ambientais. Critérios de seleção de modelos serão utilizados para a escolha do modelo mais adequado em relação aos dados observados, que permitam selecionar, como consequência, variáveis que melhor expliquem o nível de sombreamento. A forma de organização do trabalho é feita como segue:

O capítulo 2 apresentará a fundamentação teórica dos conceitos básicos da probabilidade e estatística, a incluir, definições, propriedades e proposições. Os modelos de probabilidade de interesse para esse trabalho serão definidos. Também será apresentado o modelo de regressão logística, suas propriedades e generalização, a regressão logística multinomial.

O capítulo 3 descreverá o problema da indicação do nível de sombreamento em que bovinos de corte se encontram. As variáveis envolvidas no problema serão descritas bem como a modelagem assumida.

No Capítulo 4, serão apresentadas as conclusões do trabalho.

Capítulo 2

Fundamentação Teórica

Este capítulo apresenta definições e conceitos referentes à teoria de probabilidade e estatística. O objetivo é construir embasamento teórico necessário para a definição da regressão logística multinomial, utilizada na solução do problema descrito no capítulo 3. Os conceitos aqui apresentados foram buscados em [13], [14] e [15].

2.1 Probabilidade

Esta seção apresenta a definição axiomática de probabilidade e descrição dos modelos de probabilidade binomial e multinomial. Em problemas práticos encontram-se situações que envolvem algum tipo de incerteza, que são denominados fenômenos ou experimentos aleatórios.

Definição 2.1. ***Experimentos aleatórios*** *são experimentos que mesmo repetindo o mesmo processo em condições semelhantes obtém-se resultados diferentes.*

Uma caraterística dos experimentos aleatórios é que embora não se possa dizer qual resultados em particular irá com certeza ocorrer, é possível descrever todos os possíveis resultados.

Definição 2.2. *O conjunto de todos possíveis resultados de um experimento aleatório é denominado* ***Espaço Amostral****.*

O espaço amostral é denotado por Ω. O espaço amostral Ω pode ser enumerável, finito ou infinito, caso seja possível uma correspondência biunívoca com os números naturais. Caso contrário, será não enumerável, como a reta real. Cada possível resultado do experimento aleatório, denotado por ω, é denominado ponto ou elemento de Ω. Escreve-se $\omega \in \Omega$. O conjunto vazio, denotado por $\emptyset$ é o evento sem elementos.

Dado um experimento aleatório e seu espaço amostral Ω, em geral, existe o interesse em subconjuntos de Ω, denominados eventos.

Definição 2.3. ***Evento*** *é qualquer subconjunto do espaço amostral* Ω*. Eventos são denotados por letras maiiúsculas do início do alfabeto.*

Exemplo 2.1. O lançamento de um dado honesto é um experimento aleatório. Seu espaço amostral pode ser escrito através do número de cada face de um lado: $\Omega = \{1, 2, 3, 4, 5, 6\}$. Um evento poderia ser durante o lançamento sair um número par, $A = \{2, 4, 6\}$, formando um subconjunto de Ω

Por definição, um espaço amostral e um evento carregam todas as propriedades da teoria dos conjuntos.

Definição 2.4. *Seja* Ω *um espaço amostral e dois eventos* A *e* B*,* $A, B \subset \Omega$*. Define-se como:*

Complementar de A*, denotado por* A^c*, como o evento que ocorre se e somente se* A *não ocorre, isto é:* $A^c = \{x \in \Omega; x \notin A\}$

União de A e B*, denotado por* $A \cup B$*, quando tanto* A *ou* B *ocorrem, isto é:* $A \cup B = \{w \in \Omega; w \in A \text{ ou } w \in B\}$

Interseção de A e B*, denotado por* $A \cap B$*, quando* A *e* B *ocorrem simultaneamente, isto é:* $A \cap B = \{w \in \Omega; w \in A \text{ e } w \in B\}$

As operações entre eventos permitem a determinação de novos eventos que são de interesse. Dado um espaço amostral Ω, é possível obter vários conjuntos formados por subconjuntos de Ω, denominados de classe de subconjuntos de Ω. Como consequência, faz-se necessário definir propriedades que uma classe de subconjuntos $\mathcal{F}$ deve ter de modo que as operações entre os eventos tenham sentido para o cálculo de probabilidades:

Definição 2.5. *Seja* $\mathcal{F}$ *uma coleção de subconjuntos de* Ω*. A coleção de subconjuntos* $\mathcal{F}$ *é denominada de* σ***-álgebra*** *se satisfaz as seguintes propriedades*

(i) $\Omega \in \mathcal{F}$*, isto é, o espaço amostral pertence à classe;*

(ii) se $A \in \mathcal{F}$*, então* $A^c \in \mathcal{F}$*, se um evento* A *pertence à classe, então seu complementar também pertence;*

(iii) se $A_1, A_2 ... \in \mathcal{F}$*, então* $\bigcup_{i=1}^{n} A_i \in \mathcal{F}$

Note que os items *(i)* e *(ii)* garantem o fechamento da **σ-álgebra** em relação às operações sobre conjuntos. Dessa forma sempre em que haja a atribuição de *medida de probabilidade* a um evento deve-se atribuir uma medida de probabilidade a seu complemento.

Uma classe de subconjuntos de Ω formada por todos os subconjuntos de Ω é denominada **conjunto das partes de Ω**, denotada por $\mathcal{P}(\Omega)$. Se $\mathcal{F}$ é o conjunto das partes de Ω, então $\mathcal{F}$ é um σ-álgebra de subconjuntos de Ω.

Somente sobre as classes de subconjutos $\mathcal{F}$ que satisfaçam a definição (2.5) é atribuida uma medida de probabilidade ou seja, uma σ-álgebra $\mathcal{F}$ que representa o domínio de uma função $P : \mathcal{F} \to [0, 1]$.

Definição 2.6. *Uma função P, definida na σ-álgebra de subconjuntos de Ω e com valores em $[0, 1]$ é uma medida de probabilidade se satisfaz os axiomas de Kolmogorov:*

Axioma 1. $P(\Omega) = 1$

Axioma 2. *Para qualquer evento $A \subset \Omega$, então $0 \leq P(A) \leq 1$*

Axioma 3. *Se $A_1, A_2, ... \in \mathcal{F}$ são eventos dois a dois dijuntos, isto é, $A_i \cap A_j = \emptyset$ para todos i,j com $i \neq j$, então*

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

A tripla $(\Omega, \mathcal{F}, P)$é denominada espaço de probabilidade

Exemplo 2.2. Seja $\Omega = \{\omega_1, \ldots, \omega_n\}$ um espaço amostral finito e números não negativos $p_1, \ldots, p_n$. Seja também $\mathcal{F}$ o conjunto das partes de Ω. Se $0 \leqslant p_i \leqslant 1$ e $\sum_{i=1}^{n} p_i = 1$, então para $A \in \mathcal{F}$

$$P(A) = \sum_{\omega_i \in A} p_i \tag{2.1}$$

Tomar p_i equiprovável, $p_i = \frac{1}{n}$, $i = 1, \ldots, n$, implica em que

$$P(A) = \sum_{\omega_i \in A} \frac{1}{n} = \frac{n(A)}{n} \tag{2.2}$$

onde $n(A)$ é o número de elementos de A. Dessa forma obtém-se a definição clássica de probabilidade como caso particular da definição axiomática.

2.1.1 Variável Aleatória

Em problemas práticos os resultados de um experimento aleatório levam a um número real x. Assim, a idéia de variável aleatória é associar os eventos espaço amostral Ω com intervalos do conjunto dos números reais.

Definição 2.7. *Seja E um experimento aleatório e $(\Omega, \mathcal{F}, P)$ seu espaço de probabilidade. Uma* ***variável aleatória*** *é uma função $X : \Omega \to \mathbb{R}$, que associa um intervalo real a cada evento de Ω.*

Quando uma variável aleatória X assume apenas valores em um conjunto finito ou infinito e enumerável denomina-se **variável aletória discreta**. Ao definir uma variável aleatória, um novo espaço amostral $\mathcal{X}$ é definido. Dessa forma faz-se necessária a definição da função de probabilidade, originalmente definida para o espaço amostral original Ω, que atribui uma probabilidade π para cada possível valor de X.

Definição 2.8. *A* ***função de probabilidade*** *de uma variável aleatória discreta X é uma função $f_X(x)$ que atribuí probabilidade a cada um dos possíveis valores assumidos pela variável. Isto é, sendo X uma variável aleatória discreta e $\mathcal{X} = \{x_1, x_2, \ldots x_n\}$, então*

$$f_X(x) = P(X = x) = \pi(x) = \sum_{\{w \in \Omega; X(w) = x\}} P(\{w \in \Omega; X(w) = x\}) \tag{2.3}$$

para $x \in \mathcal{X}$

A função de probabilidade de X satisfaz as seguintes propriedades:

(i) $0 \leq \pi(x) \leq 1,$ para todo $x \in \mathcal{X}$

(ii) $\sum_{x \in \mathcal{X}} \pi(x) = 1$

Dado uma variável aleatória discreta X e sua função de probabilidade $f_X(x) = P(X = x)$, existe o interesse em certas características numéricas. Essas características são conhecidas como parâmetros. A seguir definem-se os principais parâmetros para uma variável aleatória discreta: o valor esperado (esperança) e a variância.

Definição 2.9. *Seja X uma variável aleatória com $\mathcal{X} = \{x_1, \ldots, x_n\}$, e $\pi(x_i) = \pi_i$ é a função de probabilidade associada a cada ponto de $\mathcal{X}$. O* ***valor esperado*** *ou* ***esperança*** *μ é dada por:*

$$\mu = E(X) = \sum_{i=1}^{n} x_i.\pi(x_i) \tag{2.4}$$

*A **variância** σ^2 é dada por:*

$$\sigma^2 = Var(X) = E[(X - \mu)^2] \tag{2.5}$$

*A raiz quadrada da variância $\sqrt{\sigma^2}$ é chamada de **Desvio-padrão**.*

Os modelos de probabilidade, são modelos matemáticos utilizados para descrever características e comportamentos de um experimento aleatório. Neste texto os modelos probabilísticos são caracterizados em termos de seus dois principais parâmetros: Média (ou valor esperado) e Variância.

2.1.2 Modelo de Bernoulli

Esse modelo é utilizado quando o experimento aleatório admite apenas dois resultados: sucesso com probabilidade π e fracasso com probabilidade $1-\pi$. A variável aleatória X associa sucesso ao número 1 e 0 o não-sucesso, ou fracasso, $P(X = 1) = \pi$ e $P(X = 0) = 1 - \pi$, $0 < \pi < 1$.

Definição 2.10. *Seja X uma variável aleatória. X segue o modelo de Bernoulli com parâmetro π se sua função de probabilidade é dada por*

$$P(X = x|\pi) = \pi^x(1 - \pi)^{1-x},\ x = 0, 1 \tag{2.6}$$

Notação: $X \backsim Bernoulli(\pi)$

Exemplo 2.3. Considere o lançamento de um dado honesto e o interesse do pesquisador, o sucesso, em um número par como resultado do lançamento. Dessa forma o subconjunto $A = \{2, 4, 6\}$ do espaço amostral $\Omega = \{1, 2, 3, 4, 5, 6\}$ é representado quando a variável aleatória X assume valor 1, assumindo valor 0 quando o resultado do lançamento é ímpar. Pela definição clássica de probabilidade $\pi = 0.5$. Assim a probabilidade de um resultado par no lançamento de um dado honesto é dada por:

$$P(X = 1|\pi) = (0.5)^1 (1 - 0.5)^{1-1} = 0.5 \tag{2.7}$$

Proposição 2.1. Se X é uma variável aleatória com distribuição Bernoulli de parâmetro π, então, o valor esperado e a variância de X são dadas, respectivamente, por

$$E(X) = \pi \quad e \quad Var(X) = \pi(1 - \pi) \tag{2.8}$$

Demonstração. Como a X assume somente os valores 1 e 0 com probabilidades π e $1-\pi$, respectivamente, então

$$E(X) = 1\pi + 0(1-\pi) = \pi$$

$$E(X^2) = 1^2\pi + 0^2(1-\pi) = \pi$$

Logo

$$Var(X) = E(X^2) - (E(X))^2 = \pi - \pi^2 = \pi(1-\pi)$$

□

2.1.3 Modelo Binomial

Um experimento binomial é feito pela repetição de um experimento de Bernoulli n vezes de forma independente. O experimento consiste de n ensaios, cada ensaio tendo somente dois resultados, para a variável definida em (2.1.2) 0 ou 1. Cada ensaio é feito de maneira independente onde π é a probabilidade de obter um sucesso. A variável aleatória X nesse experimento descreve "número de vezes que obtém-se sucesso nos n ensaios". Obter x sucessos implica $(n-x)$ fracassos. Como o interesse não é na ordem em que eles ocorrem e sim na quantidade de vezes que eles ocorrem utiliza-se a combinação de n elementos x a x.

Definição 2.11. *A variável aleatória X segue o modelo binomial de parâmetros n e π se sua função de probabilidade é dada por:*

$$P(X = x|n,\pi) = \binom{n}{x}\pi^x(1-\pi)^{n-x},\ x = 0,1,\ldots,n \tag{2.9}$$

Notação: $X \backsim Binomial(n,\pi)$

Exemplo 2.4. Considere que o experimento realizado no exemplo (2.3) seja realizado 10 vezes, ou seja, que um dado honesto seja lançado 10 vezes. Considere que o pesquisador tenha o interesse na probabilidade em que dos 10 lançamentos 9 sejam pares. Contando com o mesmo parâmetro $\pi = 0.5$ do experimento de Bernoulli o parâmetro n é dado pelo número 10, quantidade correspondente ao número de ensaios. Assim a probabilidade para $X = 9$ é dado por:

$$P(X = 9|n,\pi) = \binom{10}{9}(0.5)^9(1-0.5)^{10-9} \approx 0.0097 \tag{2.10}$$

Proposição 2.2. Se X tem distribuição binomial de parâmetros n e π, então, o valor esperado e a variância de X são dadas, respectivamente, por:

$$E(X) = n\pi \quad e \quad Var(X) = n\pi(1-\pi) \tag{2.11}$$

Demonstração. Da definição do valor esperado (2.9),

$$\begin{aligned} E(X) &= \sum_{x=0}^{n} x \binom{n}{x} \pi^x (1-\pi)^{n-x} \\ &= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} \pi^x (1-\pi)^{n-x} \\ &= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} \pi^x (1-\pi)^{n-x} \end{aligned}$$

Tomando $k = x - 1$,

$$\begin{aligned} E(X) &= \sum_{k=0}^{n-1} \frac{n!}{k!(n-k-1)!} \pi^{k+1} (1-\pi)^{n-k-1} \\ &= n\pi \underbrace{\sum_{k=o}^{n-1} \frac{(n-1)!}{k!(n-k-1)!} \pi^k (1-\pi)^{n-k-1}}_{Binomio\ de\ Newton} \\ &= n\pi[\pi + (1-\pi)]^{n-1} \\ &= n\pi \end{aligned}$$

Analogamente, $E(X^2) = n(n-1)\pi^2 + n\pi$. Logo

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 \\ &= n(n-1)\pi^2 + n\pi - n^2\pi^2 \\ &= n^2\pi^2 - n\pi^2 + n\pi - n^2\pi^2 \\ &= n\pi - n\pi^2 \\ &= n\pi(1-\pi) \end{aligned}$$

□

2.1.4 Modelo Multinomial

Considere o caso de um experimento realizado n vezes de forma independente e produzindo resultados $a_1, a_2, \ldots, a_k$ associados respectivamente às probabilidades $\pi_1, \pi_2, \ldots, \pi_k$, com a propriedade de que $\sum \pi_i = 1$. Seja X_i o "número de vezes que obtém-se um resultado a_i nos n experimentos", $i \in \{1, \ldots, k\}$. Nesse caso é possível construir uma variável aleatória k-dimensional $\boldsymbol{X} = (X_1, \ldots, X_k)$.

Definição 2.12. *A variável aleatória $\boldsymbol{X} = (X_1, \ldots, X_k)$ segue uma distribuição multinomial com parâmetros $n = \sum_{i=1}^{n} x_i$, e $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)$ se sua função de probabilidade conjunta é*

dada por:

$$P(\boldsymbol{X} = x|\boldsymbol{\pi}) = P(X_1 = x_1, \ldots, X_k = x_k|n, \boldsymbol{\pi}) = \binom{n}{x_1, \ldots, x_k} \prod_{i=i}^{k} \pi_i^{x_i} \tag{2.12}$$

Notação: $\boldsymbol{X} \backsim Multi(n, \pi_1, \ldots, \pi_k)$

Exemplo 2.5. Considere o lançamento independente de cinco dados honestos. Considere o interesse em obter um par de "2's", um par de "4's" e um "6". O número de lançamentos implica que $n = 5$ e as probabilidades associadas aos resultados X_2, X_4 e X_6 são todas iguais a $\frac{1}{6}$. Assim a probabilidade para o evento de interesse é dada pela função de probabilidade conjunta:

$$P(X_2 = 2, X_4 = 2, X_6 = 1) = \frac{5!}{2!2!1!} \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^1 \approx 0.0038 \tag{2.13}$$

Proposição 2.3. Se a variável aleatória $\boldsymbol{X} = (X_1, \ldots, X_k)$ segue o modelo multinomial de parâmetros n, e $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)$ então $X_i \backsim Binomial(n, \pi_i)$ e

$$E(X_i) = n\pi_i \qquad Var(X_i) = n\pi_i(1 - \pi_i) \tag{2.14}$$

Demonstração. Como o experimento é repetido n vezes de modo independente e X_i o "número de vezes que obtém-se um resultado a_i"com probablidade π_i segue da definição da distribuição binomial (2.9) que $X_i \backsim Binomial(n, \pi_i)$. Por (2.2), $E(X_i) = n\pi_i$ *e* $Var(X_i) = n\pi_i(1 - \pi_i)$ □

É importante notar que escolher adequadamente o modelo de distribuição de uma variável resposta é apenas um passo na análise de dados. Na prática os parâmetros dos quais os modelos dependem são desconhecidos e dependendo da inferência estatística para sua estimação.

2.2 Estatística

O desenvolvimento de uma análise estatística utiliza dois tipos de conjuntos de dados, chamados de **população** e **amostra**. Como a população reúne todos os resultados de interesse, seu estudo em geral é impraticável ou de alto custo. Por isso o estudo de um conjunto de observações da mesma, denominada amostra, é feito para descrever o comportamento da população toda. Para uma amostra com n elementos é definido um vetor de variáveis aleatórias $\boldsymbol{X} = (X_1, \ldots, X_n)$ para representar as medidas ou valores amostrais observados [13]. Essas variáveis constituem uma **amostra aleatória** com distribuição de probabilidades $F(\theta)$, onde θ é, em geral, um parâmetro desconhecido.

Um dos objetivos da estatística é a partir de elementos $x_1, x_2, \ldots, x_n$ observados de uma amostra com distribuição $F(\theta)$ encontrar o valor mais adequado para θ. Uma função $T(\boldsymbol{X})$ que dependa apenas dos valores amostrais observados e, portanto, possa ser inteiramente calculada a partir da amostra é denominada *Estatística*. Se uma estatística contém toda a informação do parâmetro de interesse θ presente na amostra então esta estatística é dita *suficiente* para θ. O teorema (2.1), conhecido como *teorema da fatoração*, é um critério para encontrar estatísticas suficientes [13].

Teorema 2.1. *Considere Θ o conjunto onde um parâmetro θ assuma valores. Seja $\boldsymbol{X} = (X_1, \ldots, X_n)$ uma amostra aleatória com função de probabilidade $f(x_1, \ldots, x_n|\theta), \theta \in \Theta$. A estatística T é suficiente para θ, se, e somente se, existirem funções $g(t;\theta)$ e $h(x_1, \ldots, x_n)$, tais que,*

$$f(x_1, \ldots, x_n|\theta) = h(x_1, \ldots, x_n)g(t;\theta)$$

para todo $(x_1, \ldots, x_n) \in \mathbb{R}^n$ e todo $\theta \in \Theta$. A estatística T e o parâmetro θ podem ser vetores.

Considere uma amostra $\boldsymbol{X} = (X_1, \ldots, X_n)$, cuja distribuição seja Binomial ou Multinomial com parâmetro desconhecido. Pelo teorema da fatoração (2.1) é possível a obtenção de estatísticas suficientes que estimem o valor adequado para π. No entanto a pertinência dessas distribuições à família exponencial de distribuições, garante a existência de estatísticas suficientes.

2.2.1 A Família exponencial de distribuições

A distribuição de Bernoulli, a Binomial e a Multinomial fazem parte de uma família paramétrica conhecida como familia exponencial de distribuições. Essa família pode ser subdividida como uniparamétrica (apenas um parâmetro) ou multiparamétrica. Um estudo mais aprofundado sobre a familia exponencial de distribuições pode ser encontrado em [15]

A **familia exponencial uniparamétrica** é caracterizada por ter como função (de probabilidade ou densidade):

$$f(x|\theta) = h(x)\exp[\eta(\theta)t(x) - b(\theta)] \tag{2.15}$$

onde as funções $\eta(\theta), b(\theta), t(x)$ e $h(x)$ assumem valores em subconjuntos dos reais.

Pelo teorema da fatoração (2.1), considerando uma amostra aleatória $\boldsymbol{X} = (X_1, \ldots, X_n)$, $t(x)$ é uma estatística suficiente para θ.

A distribuição Binomial com índice n e parâmetro π tem como função de probabilidade:

$$P(x|\pi) = \binom{n}{x}\pi^x(1-\pi)^{n-x} = \binom{n}{x}\exp\left[x\log\left(\frac{\pi}{1-\pi}\right) + n\log(1-\pi)\right] \tag{2.16}$$

onde $\eta(\pi) = \log(\pi/1-\pi), b(\pi) = -n\log(1-\pi), t(x) = x$ e $h(x) = \binom{n}{x}$ sendo portanto parte da família exponencial uniparamétrica. A distribuição de Bernoulli, é parte dessa família considerando $n = 1$.

A familia exponêncial na forma canônica é definida por (2.15) quando $\eta(\theta)$ e $t(x)$ são iguais a função identidade, ou seja:

$$f(x|\theta) = h(x)\exp[\theta x - b(\theta)] \tag{2.17}$$

Realizando a reparametrização $\theta = log(\pi/1-\pi)$ na equação (2.16) a forma canônica da distribuição binomial é obtida pois $\eta(\theta) = \theta, t(x) = x$, são iguais a função identidade. Nesse contexto, θ pode assumir valores em toda reta real, sendo denominado **parâmetro canônico**. Esse parâmetro será utilizado para definir a função de ligação *logit* que lineariza o modelo.

A **familia exponencial multiparamétrica** de dimensão k é caracterizada por ter como função (de probabilidade ou densidade):

$$f(x|\boldsymbol{\theta}) = h(x)\exp\left[\sum_{i=1}^{k}\eta_i(\boldsymbol{\theta})t_i(x) - b(\boldsymbol{\theta})\right] \tag{2.18}$$

onde $\boldsymbol{\theta}$ é um vetor k-dimensional de parâmetros e as funções $\eta_i(\boldsymbol{\theta}), b_i(\boldsymbol{\theta}), t_i(x)$ e $h_i(x)$ têm valores em subconjuntos dos reais, generalizando a equação (2.15). Quando $\eta_i(\boldsymbol{\theta}) = \theta_i$, $i = 1, \ldots, k$ obtém-se sua forma canônica com parâmetros canônicos $\theta_1, \ldots, \theta_k$ e estatísticas canônicas $T_1(\boldsymbol{X}), \ldots, T_k(\boldsymbol{X})$. Assim (2.18) assume:

$$f(x|\boldsymbol{\theta}) = h(x)\exp\left[\sum_{i=1}^{k}\theta_i t_i(x) - b(\boldsymbol{\theta})\right] \tag{2.19}$$

Pode-se mostrar que a distribuição multinomial pertence a essa familia, isto é, que sua função de distribuição assume essa forma e identificar seu parâmetro canônico. A distribuição multinomial tem função de probabilidade

$$P(x|\boldsymbol{\pi}) = \frac{n!}{x_1!\ldots x_k!}\pi_1^{x_1}\ldots\pi_k^{x_k} \tag{2.20}$$

em que $\sum_{i=1}^{k} x_i = n$ e $\sum_{i=1}^{k}\pi_i = 1$.

A distribuição multinomial é pertencente à familia exponêncial ao tomar como parâmetro canônico $\boldsymbol{\theta} = [\log(\pi_1), \ldots, \log(\pi_k)]^T$ e a estatística $\boldsymbol{T} = (X_1, \ldots, X_k)^T$. No entanto, devido a restrição $\sum_{i=1}^{k} \pi_i = 1$, a representação mínima da família exponencial é obtida considerando $\boldsymbol{\theta} = [\log(\pi_1/\pi_k), \ldots, \log(\pi_{k-1}/\pi_k)]^T$ e $\mathbf{t} = (x_1, \ldots, x_{k-1})^T$, vetores $k-1$ dimensionais, resultando na família exponencial multiparamétrica de dimensão $k-1$

$$P(x|\boldsymbol{\theta}) = \frac{n!}{x_1! \ldots x_k!} \exp\left[\sum_{i=1}^{k-1} \theta_i x_i - b(\boldsymbol{\theta})\right], \tag{2.21}$$

onde $\theta_i = \log(\pi_i/\pi_k), i = 1, \ldots, k-1$ e $b(\boldsymbol{\theta}) = n \log\left(1 + \sum_{i=1}^{k-1} e^{\theta^i}\right)$.

O parâmetro canônico $\theta_i = \log(\pi_i/\pi_k), i = 1, \ldots, k-1$ será de especial interesse no estudo da regressão logística multinomial.

2.2.2 Método de máxima Verossimilhança

O método da máxima verossimilhança é usado na estimação de parâmetros no presente trabalho. Esse método gera estimadores com propriedades desejáveis como a consistência, convergindo para o parâmetro a medida que a amostra cresce e eficientes assintoticamente.

Definição 2.13. *Considere Θ o espaço onde um parâmetro θ assuma valores. Se $\mathbf{X} = (X_1, X_2, \ldots X_n)$ é uma amostra aleatória proveniente de uma população com função densidade de probabilidade $f(x_i|\theta)$, então, dado $\mathbf{X} = \mathbf{x}$, a função $L : \Theta \to [0, +\infty)$ definida por*

$$L(\theta; x) = \prod_{i=1}^{n} f(x_i|\theta) \tag{2.22}$$

é uma função que varia de acordo com θ denominada de ***função de verossimilhança****.*

O método de estimação de máxima verossimilhança consiste em buscar o argumento $\hat{\theta}$ que maximize a função de verrossimilhança. Em geral maximizar diretamente $L(\theta|x)$ não é uma tarefa simples. No entanto, o valor do argumento que maximiza a função de verossimilhança maximiza também o logaritmo dessa função. O logaritmo da função de verossimilhança é denominado **função log-verossimilhança**.

$$l(\theta|x) = \log\{L(\theta|x)\} = \log \prod_{i=1}^{n} f(y_i|\theta) = \sum_{i=1}^{n} \log f(y_i|\theta) \tag{2.23}$$

Maximizar a função log-verossimilhança é mais simples pois é mais fácil maximizar uma soma de termos ao invés do produto de termos. Em muitos modelos a função $L(\theta|x)$ tem um formato

côncavo e $\hat{\theta}$ é o ponto onde a derivada é igual a 0. Nesses casos o estimador de máxima verossimilhança é dada pela solução da equação

$$\frac{dl(\theta|x)}{d\theta} = 0 \tag{2.24}$$

denominada equação de verossimilhança [13].

O valor $\hat{\theta}$ é denominado estimador de máxima verossimilhança de θ. Obtido $\hat{\theta}$, é sempre necessário verificar se a segunda derivada é negativa para garantir que a solução é um ponto de máximo, dessa forma sendo necessário:

$$\left.\frac{d^2l(\theta|x)}{d\theta^2}\right|_{\theta=\hat{\theta}} < 0 \tag{2.25}$$

Denota-se $\boldsymbol{\theta}$ um vetor de parâmetros multidimensional. O estimador de máxima verossimilhança (EMV) $\hat{\boldsymbol{\theta}}$ é a solução de um conjunto de equações de verossimilhança.

2.2.3 Modelos de Regressão

Considere duas variáveis X,Y e uma amostra formada por n pares (x_i, y_i), $i = 1, \ldots, n$ dessas variáveis. A análise de regressão tem como objetivo descrever qual o comportamento da variável Y em relação à variável X cujo comportamento é conhecido. Conhecer essa relação permite obter o poder de predição de Y em função de X. Quando existe uma relação *linear* entre X e Y, diz-se que essas variáveis seguem um modelo de **regressão linear simples**. A variável Y é denominada como *variável resposta* e a variável X como *variável regressora*. Essa relação pode ser descrita como um modelo estatístico:

$$Y_i = \theta_0 + \theta_1 x_i + \varepsilon_i, i = 1, \ldots, n \tag{2.26}$$

onde ε_i é o erro, isto é, o desvio da observação em relação a média.

Nesse modelo algumas suposições são necessária:

(i) Os erros possuem média zero, ou seja, $E(\varepsilon_i) = 0$. Isso implica que $E(Y_i|x_i) = \theta_0 + \theta_1 x_i$. Ou seja, o valor observado de Y_i está próximo do valor da função de regressão com um desvio ε_i.

(ii) Os erros devem ter variância constante em torno de 0. Isto é $Var(\varepsilon_i|x_i) = \sigma^2$, para todo $i = 1, \ldots, n$.

(iii) O erro de uma observação não está correlacionado com o de outra observação, isto é, $Cov(\varepsilon_i, \varepsilon_j) = 0$, para $i \neq j$.

(iv) Em geral, os desvios seguem distribuição aproximadamente normal com $\varepsilon \backsim N(0, \sigma^2)$. Como $Y_i = \theta_0 + \theta_1 x_i + \varepsilon_i$, então $Y_i \backsim N(\theta_0 + \theta_1 x_i, \varepsilon_i)$. Logo Y_i e Y_j são independentes, para todo $i, j = 1, \ldots, n$, $i \neq j$.

O método usual para a estimação de parâmetros do modelo linear é o **método dos mínimos quadrados**. Esse método visa minimizar a soma dos quadrados dos erros obtendo estimadores ideais. Mais informações podem ser encontradas em [17].

2.2.4 Regressão Logística

O modelo de regressão logística busca modelar a situação onde a relação entre variável X e Y não é linear. Nessa situação a variável resposta dependente Y é uma variável categórica (ou qualitativa) e não de natureza contínua. Essa variável é composta apenas por um número finito, de duas ou mais categorias não permitindo o uso da modelagem linear.

Seja uma variável resposta binária, consistindo de sucesso ($Y_i = 1$) ou fracasso ($Y_i = 0$). Considere uma amostra com n unidades formada pela variável explicativa (independente) X e pela variável resposta categórica (dependente) Y, gera-se pares (x_i, y_j), $i = 1, 2, \ldots, n$ $j = 0, 1$. Assim é possível assumir $Y_i = 1$ se o evento de interesse ocorre e $Y_i = 0$ caso contrário. Pela seção (2.1.2) a variável resposta Y assume uma distribuição de Bernoulli de parâmetro π onde $P(Y_i = 1) = \pi$ e $P(Y_i = 0) = 1 - \pi$. Se a relação entre X e Y fosse linear então

$$Y_i = \theta_0 + \theta_1 x_i + \varepsilon_i \tag{2.27}$$

Aplicar as suposições do modelo linear nessa situação acarreta alguns problemas.

(i) **O valor esperado do modelo não estaria bem definido**

Uma das suposições do modelo linear é $E(\varepsilon_i) = 0$, que implica que $E(Y_i|X) = \theta_0 + \theta_1 x_i$, que não é possível pois pela proposição (2.1), $E(Y_i|X) = \pi$ é constante.

(ii) **A variância não é constante**

Como $Y_i \backsim Bernoulli(\pi)$ e pela proposição (2.1) $Var(Y_i) = \pi(1-\pi) = (\theta_0 + \theta_1 x_1)(1 - \theta_0 - \theta_1 x_1)$, ou seja, a variância de Y_i não é constante.

(iii) **Os erros não são normais**

Por (2.27) se $Y_i = 1$ então $\varepsilon_i = 1 - \theta_0 - \theta_1 x_i$. Se $Y_i = 0$ então $\varepsilon_i = -\theta_0 - \theta_1 x_i$, assumindo uma distribuição de Bernoulli e não normal.

2.2.5 Regressão Logística Binária

A regressão logística binária considera o caso onde há uma variável independente X e uma variável dependente Y, onde Y é categorizado em dois níveis.

Antes de definir a regressão logística binária uma introdução aos Modelos Lineares Generalizados faz-se necessária. Um conjunto de técnicas estatísticas normalmente estudadas e usadas separadamente podem ser formuladas de maneira unificada como uma classe de modelos de regressão. Esses modelos denominados **Modelos Lineares Generalizados** (MLG) envolvem uma variável resposta dependente univariada, variáveis explanatórias independentes e uma amostra de n observações independentes [15]. Esses MLG são usados quando uma única variável aleatória Y está associada a um conjunto de variáveis explicativas $X_0, X_1, \ldots, X_r$. Em uma amostra com n observações $(\boldsymbol{x_i}, y_i)$, onde $\boldsymbol{x_i} = (x_{i0}, \ldots, x_{ir})^T$ é o vetor coluna de variáveis explicativas, o MLG envolve os três componentes:

O **componente aleatório** do modelo é representado por um conjunto de variáveis aleatórias independentes $Y_1, \ldots, Y_n$ obtidas de uma mesma distribuição da família exponencial de distribuições com médias $\mu_1, \ldots, \mu_n$. Uma distribuição dessa familia tem expressão de probabilidade (ou densidade) dada por (2.15) ou (2.18) e, pelo teorema da fatoração (2.1), permitem a obtenção de estatísticas suficientes.

As variáveis explanatórias entram no modelo na forma de soma linear de seus efeitos formando o **componente sistemático** do modelo:

$$\eta_j = \sum_{l=o}^{r} \theta_l x_{jl} \text{ ou } \boldsymbol{\eta} = \boldsymbol{X\theta} \tag{2.28}$$

sendo $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$ a matriz do modelo, $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_r)^T$ o vetor de parâmetros desconhecidos e $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)^T$ o preditor linear.

A **função de ligação** entra no modelo ligando o valor esperado do componente aleatório ao componente sistemático, obtendo uma função $g(\mu_i) = \eta_i$, onde $g(.)$ é uma função monótona e diferenciável.

Quando $g(\mu_i)$ é a função identidade obtem-se o modelo linear. O modelo de regressão logística usa como função de ligação o **parâmetro canônico** definido na subseção (2.2.1) [14].

A regressão logística binária é um MLG. O componente sistemático é formado apenas por uma variável explanatória X. Seu componente aleatório é formado por um conjunto de variáveis independentes $Y_1, \ldots, Y_n$ que seguem uma distribuição binomial de parâmetros $n = 1$ e π. Como apresentado na subseção (2.2.1) a distribuição binomial pertence à familia exponencial.

Denotando $\pi(x) = P(Y = 1|X)$ temos que $\pi(x) = E(Y = 1|X)$. A função de ligação é o parâmetro canônico da distribuição binomial $\log(\pi/1 - \pi)$[14]. Dessa forma:

$$g(\pi(x)) = log\left[\frac{\pi(x)}{1-\pi(x)}\right] = \theta_0 + \theta_1 x \tag{2.29}$$

Essa função de ligação que lineariza o modelo é conhecida como *logit* ("**log**istic un**it**") [14].

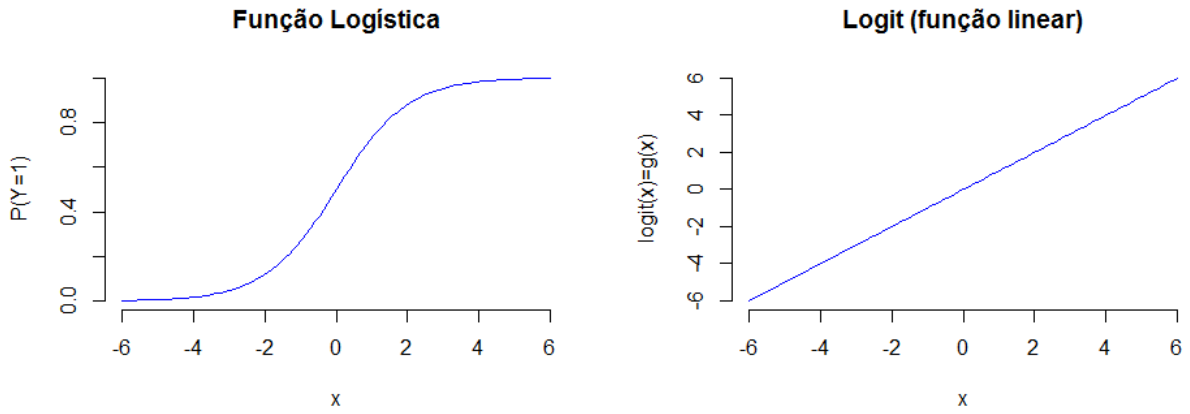

Figura 2.1: Função Logística e a relação logit

A função *logit*, conforme a figura (2.1) obtida pelo software R, é adequada para esse modelo pois é linear nos parâmetros, podendo ser contínua com valores variando em $\mathbb{R}$. A razão $(\pi/1 - \pi)$ é conhecida como razão de chances (odds ratio). A equação (2.29) permite obter uma expressão para $\pi(x)$ em função de x.

Proposição 2.4. Uma expressão para as probabilidades de um modelo de duas categorias é dada por:

$$P(Y = 1|X) = \pi(x) = \frac{e^{\theta_0+\theta_1 x}}{1 + e^{\theta_0+\theta_1 x}} \quad \text{e} \quad P(Y = 0|X) = 1 - \pi(x) = \frac{1}{1 + e^{\theta_0+\theta_1 x}} \tag{2.30}$$

Demonstração. Pela definição da transformação *logit* e propriedades dos logaritmos:

$$\begin{aligned} log\left[\frac{\pi(x)}{1-\pi(x)}\right] &= \theta_0 + \theta_1 x \\ \frac{\pi(x)}{1-\pi(x)} &= e^{\theta_0+\theta_1 x} \\ \frac{1}{1-\pi(x)} - 1 &= e^{\theta_0+\theta_1 x} \end{aligned}$$

Logo

$$\pi(x) = \frac{e^{\theta_0+\theta_1 x}}{1 + e^{\theta_0+\theta_1 x}} \quad \text{e} \quad 1 - \pi(x) = \frac{1}{1 + e^{\theta_0+\theta_1 x}} \tag{2.31}$$

□

Uma outra diferença importante entre a regressão logística e o modelo linear está na distribuição dos erros. No modelo linear assume-se que $y_i = E(y_i|x_i) + \varepsilon_i$, $i = 1, \ldots, n$ onde ε_i é o erro, isto é, o desvio de uma observação em relação a sua média. Para o modelo linear a hipótese mais comum é que ε_i em média zero e variância constante para todas as possibilidades da variável Y. Isso não é possível quando Y é de natureza binária. Assumir $y_i = E(y_i|x_i) + \varepsilon_i$ significa assumir $y_i = \pi(x_i) + \varepsilon_i$, implicando que a variável aleatória ε pode assumir apenas duas possibilidades. Se $y_i = 1$ então $\varepsilon_i = 1 - \pi_i$ com probabilidade $\pi(x)$. Se $y_i = 0$ então $\varepsilon_i = -\pi_i$ com probabilidade $1 - \pi(x)$.

Proposição 2.5. A variável aleatória ε tem distribuição Bernoulli com média zero e variância $\pi(x)[1 - \pi(x)]$

Demonstração. O valor esperado de ε dado $X = x$ é dado por

$$E(\varepsilon|x) = \sum_{j=0}^{1} \varepsilon_j P(\varepsilon = \varepsilon_j) = -\pi(x)[1 - \pi(x)] + [1 - \pi(x)]\pi(x) = 0 \tag{2.32}$$

A variância de ε é obtida por

$$\begin{aligned} Var(\varepsilon|x) = \sum_{j=0}^{1} \varepsilon_j^2 P(\varepsilon = \varepsilon_j) &= [-\pi(x)]^2(1 - \pi(x)) + [1 - \pi(x)]^2\pi(x) \\ &= \pi(x)(1 - \pi(x)[(1 - \pi(x)) + \pi(x)] \\ &= \pi(x)(1 - \pi(x)) \end{aligned} \tag{2.33}$$

□

Assim os erros ε_i seguem as seguintes suposições, para todo $i, l = 1, \ldots, n$

$$\begin{aligned} &\text{(i) } E(\varepsilon_i|x_i) = 0 \\ &\text{(ii) } Var(\varepsilon_i|x_i) = \pi(x_i)[1 - \pi(x_i)] \\ &\text{(iii) } Cov(\varepsilon_i, \varepsilon_l) = 0, \text{ se } i \neq l \end{aligned} \tag{2.34}$$

Para ajustar esse modelo a um conjunto de dados é necessário estimar os valores de θ_0 e θ_1, parâmetros desconhecidos. O método de estimação para esses parâmetros será o de máxima verossimilhança.

Proposição 2.6. Seja $\boldsymbol{\theta} = (\theta_0, \theta_1)$ um vetor de parâmetros associada à probabilidade $P(Y_i = 1|x_i) = \pi(x_i)$. Então, o estimador, pelo método de máxima verossimilhança de $\boldsymbol{\theta}$, denotado por

$\hat{\boldsymbol{\theta}}$, é a solução das equações de verossimilhança

$$\sum_{i=1}^{n}[y_i - \pi(x_i)] = 0 \tag{2.35}$$

$$\sum_{i=1}^{n} x_i[y_i - \pi(x_i)] = 0 \tag{2.36}$$

Demonstração. Como $Y_i \backsim Bernoulli(\pi(x_i))$ sua função probabilidade é dada por

$$P\left(Y_i = y_i|\pi(x_i)\right) = \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i} \tag{2.37}$$

onde $y_i \in \{0, 1\}$ e $i = 1, \ldots, n$. Assumindo que as observações são independentes, a função de verossimilhança para $\boldsymbol{\theta}$ é

$$L(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) = \prod_{i=1}^{n} \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i} \tag{2.38}$$

Aplicando o logaritmo em (2.38):

$$\begin{aligned} l(\boldsymbol{\theta}) &= \log\left[\prod_{i=1}^{n} \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i}\right] \\ &= \sum_{i=1}^{n}[y_i \log \pi(x_i) + (1 - y_i)\log(1 - \pi(x_i))] \end{aligned} \tag{2.39}$$

Derivando $l(\boldsymbol{\theta})$ em relação à θ_0 e θ_1:

$$\begin{aligned} \frac{dl(\boldsymbol{\theta})}{d\theta_0} &= \sum_{i=1}^{n}[y_i - \pi(x_i)] = 0 \\ \frac{dl(\boldsymbol{\theta})}{d\theta_1} &= \sum_{i=1}^{n} x_i[y_i - \pi(x_i)] = 0 \end{aligned} \tag{2.40}$$

□

As equações de (2.40) não são lineares nos parâmetros e têm que ser resolvidas numericamente por processos iterativos. O processo utilizado em vários softwares estatísticos como o R é o **Método Escore de Fisher**. Esse processo iterativo é uma adaptação do método de Newton-Raphson que envolve substituir a matriz das derivadas parciais, a matriz de informação observada, pela matriz de informação $\boldsymbol{K}$ de valores esperados cujo elemento (j, k) é [15]:

$$K_{j,k} = -E\left(\frac{\partial^2 l(\boldsymbol{\theta}|x)}{\partial\theta_j \partial\theta_k}\right) \tag{2.41}$$

Considerando que se deseja obter a solução do sistema de equações $\boldsymbol{U} = \boldsymbol{U}(\boldsymbol{\theta}) = \partial l(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \mathbf{0}$ e, usando a versão multivariada do método Newton-Raphson, tem-se

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} + (\boldsymbol{K}^{(m)})^{-1}\boldsymbol{U}^{(m)}, \tag{2.42}$$

sendo $\boldsymbol{\theta}^{(m)}$ e $\boldsymbol{\theta}^{(m+1)}$ os vetores de parâmetros estimados nos passos m e $(m+1)$, respectivamente, $\boldsymbol{U}^{(m)}$ o vetor escore avaliado no passo m, e $(\boldsymbol{K}^{(m)})^{-1}$ a inversa matriz de informação, com elementos $\dfrac{-\partial^2 l(\boldsymbol{\theta})}{\partial\theta_j \partial\theta_k}$, avaliada no passo m.

O método usual para iniciar o processo iterativo é especificar uma estimativa inicial e, sucessivamente, alterá-la até que a convergência seja alcançada. Define-se $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(m+1)}$.

Os métodos de inferência nos modelos de regressão logística baseiam-se, essencialmente, na teoria de máxima verossimilhança [17]. A avaliação da significância das variáveis é feita pela comparação entre um modelo saturado e um modelo ajustado (um modelo é dito saturado se contém tantos parâmetros quantos dados observados), sendo baseada na estatística D definida a seguir:

$$D = -2\log\left[\frac{\text{verossimilhança do modelo ajustado}}{\text{verossimilhança do modelo saturado}}\right] \tag{2.43}$$

A estatística D é chamada *deviance* e auxilia na comparação entre os valores observados e preditos. Para avaliar a significância de uma variável independente compara-se o valor de D com e sem a variável, reprsentados por D_M e D_0 respecivamente. A comparação é feita através da estatística G definida por:

$$G = D_0 - D_M \tag{2.44}$$

podendo também ser escrita como

$$G = -2\log\left[\frac{\text{verossimilhança sem a variável}}{\text{verossimilhança com a variável}}\right] \tag{2.45}$$

Avaliando a hipótese em que θ_1 é igual a zero, é conhecido que a estatística G tenha uma distribução assintótica qui-quadrado ($\mathcal{X}^2$) com 1 grau de liberdade [14]. Esse fato motiva o seguinte teste de significância:

Teste de razão de verossimilhança: No modelo de regressão logística binária, o teste de razão de verossimilhança de tamanho α é dado pela expressão

$$\mathcal{H}_0 : \theta_1 = 0 \; vs \; \mathcal{H}_1 : \theta_1 \neq 0 \tag{2.46}$$

consiste em rejeitar $\mathcal{H}_0$ se $\mathbb{P}[\mathcal{X}_1^2 > G] < \alpha$.

Outros testes, como o Teste de Wald, são estatísticamente equivalentes ao teste de razão de verossimilhança. O Teste de Wald é baseado na estatística W sob as mesmas hipóteses de (2.46) [16], permitindo o teste de significância para estimadores separadamente, além de servir

de base para a construção de intervalos de confiança. A estatística W é definida como

$$W = \frac{\hat{\theta}_1}{\widehat{SE}(\hat{\theta}_1)} \tag{2.47}$$

que consiste em comparar o estimador de θ_1 com seu erro padrão $\widehat{SE}(\hat{\theta}_1)$. Sob a hipótese que $\theta_1 = 0$, W possui distribuição normal [14].

Teste de Wald: No modelo de regressão logística binária, o teste de Wald de tamanho α é dado pela expressão

$$\mathcal{H}_0 : \theta_1 = 0 \; vs \; \mathcal{H}_1 : \theta_1 \neq 0 \tag{2.48}$$

consiste em rejeitar $\mathcal{H}_0$ se $\mathbb{P}(z > W) < \alpha$.

Proposição 2.7. O intervalo de $100(1-\alpha)\%$ de confiança para θ_0 é dado por

$$[\hat{\theta}_0 - z_{\frac{\alpha}{2}}\widehat{SE}(\hat{\theta}_0), \hat{\theta}_0 + z_{\frac{\alpha}{2}}\widehat{SE}(\hat{\theta}_0)] \tag{2.49}$$

onde $z_{\frac{\alpha}{2}}$ é o quantil de uma distribuição normal padrão dado por $\mathbb{P}(z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$.

Proposição 2.8. O intervalo de $100(1-\alpha)\%$ de confiança para θ_1 é dado por

$$[\hat{\theta}_1 - z_{\frac{\alpha}{2}}\widehat{SE}(\hat{\theta}_1), \hat{\theta}_0 + z_{\frac{\alpha}{2}}\widehat{SE}(\hat{\theta}_1)] \tag{2.50}$$

onde $z_{\frac{\alpha}{2}}$ é o quantil de uma distribuição normal padrão dado por $\mathbb{P}(z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$.

2.2.6 Regressão Logistica Multinomial

O modelo de regressão apresentado na seção (2.2.5) consiste de uma variável aleatória binária assumindo apenas valores zero ou um. Esse modelo usa como função de ligação a função *logit* de $Y = 1$ contra $Y = 0$. O modelo de regressão logística binária pode ser estendido de modo que exista mais de uma variável independente e a variável resposta dependente, de natureza categórica, apresente mais de dois níveis, denominada **regressão logística multinomial.**

A regressão logística multinomial é um MLG. Sua variável independente $\boldsymbol{x}$ uma coleção de $r+1$ variáveis independentes $(x_0, x_1, \ldots, x_r)$ com $X_0 = 1$ forma o **componente sistemático** do modelo na forma de uma soma linear.

A variável resposta Y é uma variável de natureza categórica que pode assumir valores de $1, \ldots, k$, tendo a distribuição multinomial como distribuição de probabilidades. Essa variável forma o **componente aleatório** do modelo. Seu parâmetro canônico apresentado na seção (2.2.1) é igual a $\theta_j = \log(\pi_j/\pi_k), j = 1, \ldots, k-1$.

Dado o componente sistemático e um componente aleatório defini-se uma função de ligação adequada, isto é uma função g que associe o valor esperado de Y à uma estrutura linear de cada variável independente. Para as k respostas de Y, associada aos valores assumidos pela variável de $\boldsymbol{x} = (x_0, x_1, \ldots, x_r)$, a função de ligação g assume:

$$\begin{aligned} g_1(\boldsymbol{x}) &= \theta_{10}x_0 + \theta_{11}x_1 + \ldots + \theta_{1r}x_r \\ g_2(\boldsymbol{x}) &= \theta_{20}x_0 + \theta_{21}x_1 + \ldots + \theta_{2r}x_r \\ &\vdots \\ g_k(\boldsymbol{x}) &= \theta_{k0}x_0 + \theta_{k1}x_1 + \ldots + \theta_{kr}x_r \end{aligned} \tag{2.51}$$

De forma análoga à regressão logística binária, é tomado o parâmetro canônico da distribuição multinomial como função de ligação g [14]. Uma extensão da função *logit* é obtida:

$$g_j(\boldsymbol{x}) = \log\left(\frac{\pi_j(\boldsymbol{x})}{\pi_k(\boldsymbol{x})}\right), j = 1, \ldots, k \tag{2.52}$$

onde $\pi_j(\boldsymbol{x}) = P(Y = y_j|\boldsymbol{x})$. Assim,

$$\begin{aligned} \log\left(\frac{P(Y=1|\boldsymbol{x})}{P(Y=k|\boldsymbol{x})}\right) &= \theta_{10}x_0 + \theta_{11}x_1 + \ldots + \theta_{1r}x_r \\ \log\left(\frac{P(Y=2|\boldsymbol{x})}{P(Y=k|\boldsymbol{x})}\right) &= \theta_{20}x_0 + \theta_{21}x_1 + \ldots + \theta_{2r}x_r \\ &\vdots \\ \log\left(\frac{P(Y=k|\boldsymbol{x})}{P(Y=k|\boldsymbol{x})}\right) &= \theta_{k0}x_0 + \theta_{k1}x_1 + \ldots + \theta_{kr}x_r = 0 \end{aligned} \tag{2.53}$$

A classe k é denominada *categoria de referência* ou categoria base e é tomado sem perda de generalidade. A variável $Y = k$ normalmente é o valor mais comum da amostra e serve para a comparação entre as respostas $1, \ldots, k-1$ e a resposta k. A propriedade dos logaritmos permitem tomar qualquer valor $m \in \{1, 2, \ldots, k-1\}$ como categoria de referencia pois:

$$\log\left(\frac{\pi_j}{\pi_m}\right) = \log\left(\frac{\pi_j}{\pi_k}\right) - \log\left(\frac{\pi_m}{\pi_k}\right) \tag{2.54}$$

A relação $\dfrac{\pi_j}{\pi_k}$ é uma extensão da razão de chances da regressão logística binária. As equações (2.53) permitem obter expressões para π_j em função de $\boldsymbol{x}$.

Proposição 2.9. Uma expressão para as probabilidades de um modelo com k categorias é dada por

$$P(Y = j|\boldsymbol{x}) = \pi_j(\boldsymbol{x}) = \frac{exp[g_j(\boldsymbol{x})]}{\sum_{j=1}^{k} exp[g_j(\boldsymbol{x})]}, j = 1, \ldots, k \tag{2.55}$$

onde $g_j(\boldsymbol{x}) = \theta_{j0}x_0 + \theta_{j1}x_1 + \ldots + \theta_{jr}x_r$ uma função dependente de $\boldsymbol{x}$ e $g_k = 0$.

Demonstração. Pela definição da transformação *logit*

$$\log\left(\frac{\pi_j(\boldsymbol{x})}{\pi_k(\boldsymbol{x})}\right) = g_j(\boldsymbol{x}), j = 1, \ldots, k$$

usando propriedade dos logaritmos

$$\frac{\pi_j(\boldsymbol{x})}{\pi_k(\boldsymbol{x})} = exp[g_j(\boldsymbol{x})]$$

$$\pi_j(\boldsymbol{x}) = exp[g_j(\boldsymbol{x})]\pi_k(\boldsymbol{x}) \tag{2.56}$$

Pela propriedade da probabilidade total a soma das k probabilidades totalizam um, isto é,

$$\sum_{j=1}^{k} \pi_j(\boldsymbol{x}) = 1 \tag{2.57}$$

Substituindo (2.56) em (2.57), obtem-se:

$$\sum_{j=1}^{k} \exp[g_j(\boldsymbol{x})]\pi_k(\boldsymbol{x}) = 1$$

A propriedade de somatório leva a

$$\pi_k(\boldsymbol{x}) = \frac{1}{\sum_{j=1}^{k} exp[g_j(\boldsymbol{x})]} \tag{2.58}$$

e, substituindo esse resultado em (2.56)

$$\pi_j(\boldsymbol{x}) = \frac{exp[g_j(\boldsymbol{x})]}{\sum_{j=1}^{k} exp[g_j(\boldsymbol{x})]} \tag{2.59}$$

□

Considere uma amostra $\boldsymbol{x_i} = (x_{i0}, \ldots, x_{ir})$ para $i = 1, \ldots, n$ com n observações permitindo a obtenção de n pares de observações $(\boldsymbol{x_i}, y_i)$ com $i = 1, \ldots, n$. Como Y_i é de natureza multinomial, com k categorias, pode-se escrever cada expressão da amostra como $y_i = E(Y_i|\boldsymbol{x_i}) + \varepsilon_i$, isto é $y_i = \pi(\boldsymbol{x_i}) + \varepsilon_i$. Assim a variável ε pode assumir k valores pois y_i assume k valores. Se $y_i = j$ então $\varepsilon_i = j - \pi(\boldsymbol{x_i})$ com probabilidade $P(Y = j|\boldsymbol{x_i})$, onde $j = 1, \ldots, k$ [17].

Proposição 2.10. A variável ε tem distribuição multinomial com média zero e variância igual a da variável Y.

Demonstração. A esperança de ε dado $\boldsymbol{x_i}$ é dada por

$$\begin{aligned} E(\varepsilon|\boldsymbol{x_i}) &= \sum_{j=1}^{k} \varepsilon_j P(\varepsilon = \varepsilon_j|\boldsymbol{x_i}) \\ &= (1-\pi(\boldsymbol{x_i}))P(Y=1|\boldsymbol{x_i}) + \ldots + (k-\pi(\boldsymbol{x_i}))P(Y=k|\boldsymbol{x_i}) \\ &= \sum_{j=1}^{k} jP(Y=j|\boldsymbol{x_i}) - \pi(\boldsymbol{x_i}) \sum_{j=1}^{k} P(Y=j|\boldsymbol{x_i}) \\ &= E(Y|\boldsymbol{x_i}) - \pi(\boldsymbol{x_i}) = \pi(\boldsymbol{x_i}) - \pi(\boldsymbol{x_i}) = 0 \end{aligned} \tag{2.60}$$

A variância de ε é obtida por:

$$\begin{aligned} Var(\varepsilon|\boldsymbol{x_i}) &= \sum_{j=1}^{k} \varepsilon_j^2 P(\varepsilon = \varepsilon_j|\boldsymbol{x_i}) = \sum_{j=1}^{k} (j-\pi(\boldsymbol{x_i})^2 P(Y=j|\boldsymbol{x_i}) \\ &= \sum_{j=1}^{k} (j^2 - 2j\pi(\boldsymbol{x_i}) + \pi^2(\boldsymbol{x_i}))P(Y=j|\boldsymbol{x_i}) \\ &= E(Y^2|\boldsymbol{x_i}) - 2\pi(\boldsymbol{x_i})E(Y|\boldsymbol{x_i}) + \pi^2(\boldsymbol{x_i}) \\ &= E(Y^2|\boldsymbol{x_i}) - [E(Y|\boldsymbol{x_i})]^2 = Var(Y|\boldsymbol{x_i}) \end{aligned} \tag{2.61}$$

□

Considerando uma amostra com n observações independentes, os erros ε_i, seguem as seguintes suposições, para todo $i, l = 1, \ldots, n$

$$\begin{aligned} &\text{(i) } E(\varepsilon_i|x_i) = 0 \\ &\text{(ii) } Var(\varepsilon_i|x_i) = Var(Y_i|x_i) \\ &\text{(iii) } Cov(\varepsilon_i, \varepsilon_l) = 0, \text{ se } i \neq l \end{aligned} \tag{2.62}$$

Considere $\boldsymbol{\theta}$ o vetor de parâmetros relacionados à probabilidade $P(Y_j = y_j|\boldsymbol{x_i})$, $i = 1, \ldots, n$ e $j = 1, \ldots, k$. O método de estimação para tais parâmetros é o método de máxima verossimilhança. No entanto, para o cálculo da função de verossimilhança é necessário o uso de uma variável latente $\boldsymbol{Z} = (z_{i1}, z_{i2}, \ldots, z_{ik})$. Essa variável assume valor $z_{ij} = 1$, se $y_i = j$ e $z_{ij} = 0$ caso contrário, para $j = 1, \ldots, k$.

Proposição 2.11. Seja $\boldsymbol{\theta}$ o vetor de parâmetros relacionados à probabilidade $P(Y = j|\boldsymbol{x_i})$, $i = 1, \ldots, n$ e $j = 1, \ldots, k-1$. Considere a variável latente $\boldsymbol{Z}$ dependente de Y onde $z_{ij} = 1$, se $y_i = j$ e $z_{ij} = 0$ caso contrário. O estimador de máxima verossimilhança de $\boldsymbol{\theta}$, denotado por $\hat{\boldsymbol{\theta}}$, é a solução das equações de verossimilhança:

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_{jl}} = \sum_{i=1}^{n} x_{il}(z_{ij} - \pi_{ij}) = 0 \tag{2.63}$$

para $j = 1, \ldots, k-1$, $l = 0, \ldots, r$ e $\pi_{ij} = \pi_j(\boldsymbol{x_i})$, com $x_{i0} = 1$ para todo i.

Demonstração. Como as variáveis latentes z_{ij} tem distribuição binomial de parâmetros $n = 1$ e π_{ij}, a variável $\boldsymbol{Z}$ tem distribuição multinomial cujo núcleo de sua função probabilidade é:

$$P(Z = z|\boldsymbol{x_i}) \propto \pi_{i1}^{z_{i1}}(\boldsymbol{x_i}) \ldots \pi_{ik}^{z_{ik}}(\boldsymbol{x_i}) \tag{2.64}$$

Da definição da função de verossimilhança:

$$\begin{aligned} L(\boldsymbol{\theta}|Y, \boldsymbol{x}) &\propto \prod_{i=1}^{n} [\pi_1^{z_{i1}}(\boldsymbol{x_i}) \ldots \pi_k^{z_{ik}}(\boldsymbol{x_i})] \\ l(\boldsymbol{\theta}) &\propto \log \{L(\boldsymbol{\theta}|Y, \boldsymbol{x})\} = \log \left\{ \prod_{i=1}^{n} [\pi_1^{z_{i1}}(\boldsymbol{x_i}) \ldots \pi_k^{z_{ik}}(\boldsymbol{x_i})] \right\} \\ l(\boldsymbol{\theta}) &\propto \sum_{i=1}^{n} [z_{i1} \log(\pi_1(\boldsymbol{x_i})) + \ldots + z_{ik} \log(\pi_k(\boldsymbol{x_i}))] \end{aligned} \tag{2.65}$$

Substituindo $\pi(\boldsymbol{x_i})$ pela expressão da proposição (2.9):

$$l(\boldsymbol{\theta}) \propto \sum_{i=1}^{n} \left\{ \sum_{j=1}^{k} z_{ij} g_j(\boldsymbol{x_i}) - \log \left[\sum_{j=1}^{k} \exp[g_j(\boldsymbol{x_i})] \right] \right\} \tag{2.66}$$

As $(k-1)(r+1)$ equações de verossimilhança são obtidas através das derivadas parciais de (2.66) em relação aos $(k-1)(r+1)$ parâmetros a serem estimados. A solução do sistema de equações de verossimilhança é o EMV $\hat{\boldsymbol{\theta}}$. □

As equações de verossimilhança são não lineares nos parâmetros e têm que ser resolvidas numericamente por processos iterativos. O método iterativo usado é o Método Escore de Fisher apresentado em (2.42) [17].

A significância dos $(k-1)(r+1)$ coeficientes é medida pelo uso da estatística G definido em (2.45), assumindo distribuição qui-quadrado com graus de liberdade equivalente ao número de parâmetros da hipótese nula. O teste é realizado com um subconjunto $\boldsymbol{\theta_1}$ de $\boldsymbol{\theta}$ para testar uma ou mais variáveis.

Teste de razão de verossimilhança: Caso Multinomial: No modelo de regressão logística multinomial, o teste de razão de verossimilhança de tamanho α para um vetor de parâmetros $\boldsymbol{\theta_1} \in \boldsymbol{\theta}$, $\boldsymbol{\theta_1}$ com q variáveis e de tamanho $q(k-1)$ parâmetros é dado por

$$\mathcal{H}_0 : \boldsymbol{\theta_1} = \boldsymbol{b_1} \; vs \; \mathcal{H}_1 : \boldsymbol{\theta_1} \neq \boldsymbol{b_1}, \text{ para } \boldsymbol{b_1} \in \mathbb{M}_{(k-1)\times(q)} \tag{2.67}$$

consiste em rejeitar $\mathcal{H}_0$ se $\mathbb{P}[\mathcal{X}^2_{q(k-1)} > G] < \alpha$, onde $\mathbb{M}_{(k-1)\times(q)}$ é o conjunto de todas as matrizes de dimensão $(k-1) \times (q)$.

A extensão da estatística W para o teste de Wald é obtida pelo uso da **matriz de covariância** de $\hat{\boldsymbol{\theta}}_{\mathbf{1}}$ cujo valor genérico foi definido em (2.41). Assim a estatística W estendida ao caso multinomial é [15]:

$$W = (\hat{\boldsymbol{\theta}}_{\mathbf{1}} - \boldsymbol{b}_{\mathbf{1}})^T \widehat{cov}(\hat{\boldsymbol{\theta}}_{\mathbf{1}})^{-1} (\hat{\boldsymbol{\theta}}_{\mathbf{1}} - \boldsymbol{b}_{\mathbf{1}}) \tag{2.68}$$

que sob a hipótese nula tem distribuição qui-quadrado com $q(k-1)$ graus de liberdade.

Teste de Wald: Caso Multinomial: No modelo de regressão logística binária, o teste de Wald de tamanho α para um vetor de parâmetros $\boldsymbol{\theta}_{\mathbf{1}} \in \boldsymbol{\theta}$, $\boldsymbol{\theta}_{\mathbf{1}}$ com q variáveis e de tamanho $q(k-1)$ parâmetros é dado pela expressão

$$\mathcal{H}_0 : \boldsymbol{\theta}_{\mathbf{1}} = \boldsymbol{b}_{\mathbf{1}} \; vs \; \mathcal{H}_1 : \boldsymbol{\theta}_{\mathbf{1}} \neq \boldsymbol{b}_{\mathbf{1}}, \text{ para } \boldsymbol{b}_{\mathbf{1}} \in \mathbb{M}_{(k-1)\times(q)} \tag{2.69}$$

consiste em rejeitar $\mathcal{H}_0$ se $\mathbb{P}[\mathcal{X}^2_{q(k-1)} > W] < \alpha$.

Proposição 2.12. O intervalo a $100(1-\alpha)\%$ de confiança para θ_{il} é dado por

$$[\hat{\theta}_{il} - z_{\frac{\alpha}{2}} \widehat{SE}(\hat{\theta}_{il}), \hat{\theta}_{il} + z_{\frac{\alpha}{2}} \widehat{SE}(\hat{\theta}_{il})] \tag{2.70}$$

onde $z_{\frac{\alpha}{2}}$ é o quantil de uma distribuição normal padrão dado por $\mathbb{P}(z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$.

Estatísticas Pseudo R Quadrado

As estatísticas Pseudo-R^2 são usados para medir o quão bem um modelo se ajusta ao conjunto de dados. As medidas R^2 indicam medidas do poder de previsão de um modelo, quanto maior seu valor melhor o poder de previsão. Das estatísticas Pseudo R Quadrado, três são de interesse: A *Cox and Snell*, a *McFadden* e a *Nagelkerke*.

Seja L_0 o valor da função de verossimilhança de um modelo sem preditores e L_M a verossimilhança do modelo a ser estimado. A estatística **McFadden** R^2 é:

$$R^2_{McF} = 1 - log(L_M)/log(L_0) \tag{2.71}$$

Denotando n o tamanho da amostra, a estatística **Cox and Snell** R^2 [18] é:

$$R^2_{C\&S} = 1 - [(L_0)/(L_M)]^{2/n} \tag{2.72}$$

A estatística **Nagelkerke** [19] é uma extensão da *Cox and Snell* e envolve sua divisão pelo seu máximo $1 - (L_0)^{2/n}$. A expressão para essa estatística é

$$R^2_{Nag} = \frac{1 - [(L_0)/(L_M)]^{2/n}}{1 - (L_0)^{2/n}} \tag{2.73}$$

Capítulo 3

Modelo Estatístico para Classificação do Nível de Sombreamento

Em regiões tropicais, o estresse térmico é um fator importante para a criação de sistemas produtivos de bovinos de corte que buscam a máxima eficiência [20]. A umidade do ar e temperatura, de acordo com [21], são os maiores responsáveis pelo conforto térmico animal e, a partir destas, é possível o desenvolvimento de índices para avaliar o estresse térmico. De acordo com [22] uma alternativa para evitar o estresse térmico é o uso do sombreamento que diminui a incidência da radiação solar sobre o animal. Variáveis como a temperatura e umidade do ar, luminosidade e radiação UV são importantes para avaliar o nível de sombreamento do bovino e identificar em quais situações o animal sofre em detrimento do estresse térmico [23].

O Laboratório de Sistemas Computacionais de Alto Desempenho (LSCAD) da Universidade Federal de Mato Grosso do Sul, desenvolveu um software que visa obter e analisar dados ambientais. A solução de hardware e software, em conjunto, busca identificar as variações de temperatura na pele do animal, bem como a indicação da sua busca por sombra. O sistema proposto por [23] é capaz de adquirir, processar, armazenar e analisar dados ambientais e indicar com base nesses dados o estresse térmico desses animais. A captação dos dados ambientais é realizada através de sensores que captam a temperatura e luminosidade do ambiente, a umidade relativa do ar e a radiação Ultra-Violeta ligados por uma plataforma eletrônica. Esses quatro dados são armazenados em forma de variáveis em um equipamento servidor onde é possível sua análise. [23] apresenta o uso de algoritmos, baseados na teoria da lógica difusa ou *Fuzzy*, para determinação do sombreamento a partir dos valores de entrada dos sensores. A lógica *fuzzy* admite o conceito de verdade parcial, consistindo de 11 valores possíveis entre 0 e 1.

O algoritmo *fuzzy* utiliza os dados ambientais (Luz, Radiação UV, Temperatura, Humidade) gerando *fuzzysets* que classificam faixas de valores de cada variável como Alta, Média ou Baixa. Ao final um conjunto de 81 regras de validação gera o resultado de saída, denominada estado. O algoritmo considera o nível de sombreamento entre 0 e 1, sendo 0 a ausência de sombra, chamado Sol, 1 a presença de sombra densa e um estado intermediário, denominado nublado.

Esse trabalho apresentará critérios de seleção de modelos estatísticos que façam uso dos dados ambientais e que possuam como variável resposta o nível de sombreamento categorizado como sol, nublado ou sombra.

3.1 Descrição do experimento

A plataforma desenvolvida pelo LSCAD, que consiste de sensores de captação de dados ambientais, é ilustrada na figura (3.1).

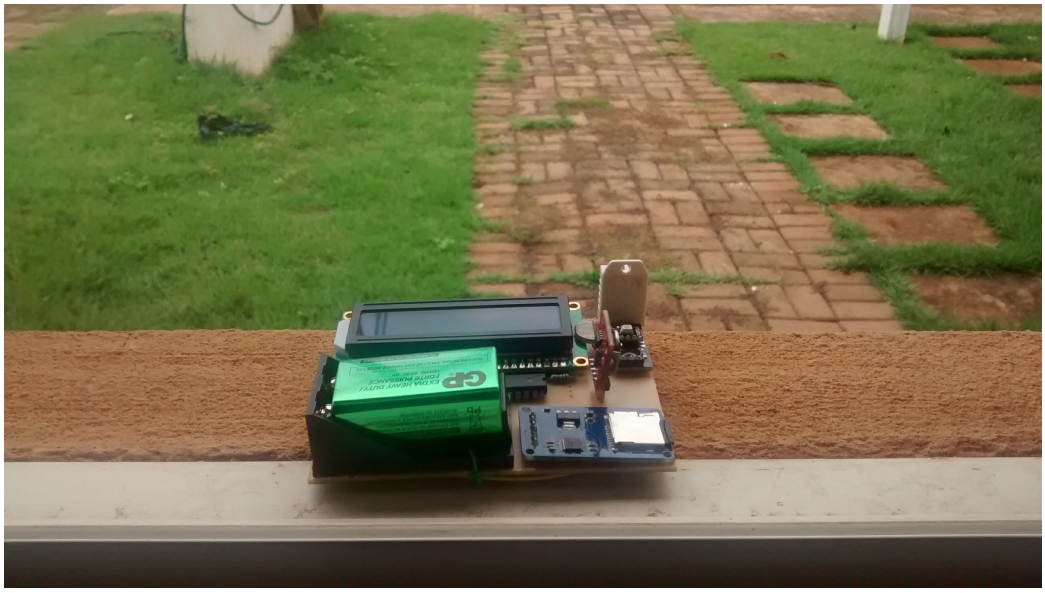

Figura 3.1: Plataforma de captação de dados ambientais

De acordo com [23] a plataforma de captação utilizada nesse trabalho contém três sensores. O sensor DHT22, ilustrado na figura (3.2), é responsável pela captação da umidade relativa do ar e temperatura ambiente. Seu alcance de detecção varia entre 5% a 99% para umidade relativa do ar, e de -40°C a 125°C para temperatura ambiente, com precisão de 2% para umidade e 0,5°C para temperatura.

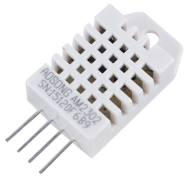

Figura 3.2: Sensor DHT22

O resistor dependente de luz ou LDR, ilustrado na figura (3.3), é responsável pela captação dos valores de luz ambiente. Seu alcance varia entre 0 e 1023, onde 0 representa resistência total, ou seja, iluminação máxima e 1023 a ausência de luz.

Figura 3.3: Sensor LDR

O sensor responsável pela captação de raios UltraVioletas (UV) é o sensor UVM-30A, ilustrado na figura (3.4). Seu alcance varia entre 0 e 11 onde 0 representa radiação baixa e 11 extrema.

Figura 3.4: Sensor de Ultravioleta UVM30A

Utilizando as variáveis explicativas obtidas dos sensores (*luminosidade*; *radiação UV*; *temperatura* e *Umidade*) propõe-se um modelo que possua como variável resposta o nível de sombreamento no instante da observação, classificada como *Sol*, *Sombra* ou *Nublado*. A análise de dados foi realizada através do software estatístico R [12] e o pacote VGAM [24]. As variáveis usadas no modelo encontram-se na tabela (3.1).

Tabela 3.1: Variáveis utilizadas para indicação do nível de sombreamento

Variável	**Descrição**	**Natureza da Variável**	**Valores assumidos**
Y	Indicador de estado de estresse térmico	Categórica	1- Sombra 2-Nublado 3-Sol
X_1	Luminosidade Natural	Discreta	1023 a 0 (Decrescente)
X_2	Radiação Ultravioleta UV	Discreta	0 a 11 (Crescente)
X_3	Temperatura em graus Celsius ($^{\circ}C$)	Contínua	-40 a 125 (Crescente)
X_4	Umidade relativa do ar	Contínua	5% a 99%

O conjunto de dados inicial é formado por 505 observações, uma leitura por minuto, totalizando oito horas e vinte e cinco minutos de coleta. Para cada leitura foi registrado o nível de sombreamento no momento da coleta agregando ao conjunto de dados inicial uma coluna denominado estado. Dessa 505 observações 49% foram leituras sob o estado "Sol", 30% leituras sob o estado "Nublado" e 21% leituras sob o estado "Sombra". A tabela (3.2) demonstra como os dados iniciais foram organizados, contendo 25 das 505 observações.

Tabela 3.2: Organização dos dados observados

Horário	X_1	X_2	X_3	X_4	Y
08/04/2017	**Luminosidade**	**Radiação UV**	**Temperatura**	**Umidade**	**Estado observado**
10:17	4	4	34.20	52.60	Nublado
10:18	3	5	33.40	52.20	Nublado
10:19	3	7	33.10	53.20	Sol
10:20	3	7	33.70	52.60	Sol
10:21	3	7	33.40	51.80	Sol
10:22	2	7	33.00	53.00	Sol
10:23	2	7	33.00	53.10	Sol
10:24	2	8	32.40	53.40	Sol
10:25	2	8	32.10	53.60	Sol
10:26	2	8	33.30	54.30	Sol
10:27	3	8	34.00	53.30	Sol
10:28	3	8	33.80	52.30	Sol
10:29	2	8	34.50	51.80	Sol
10:30	2	8	33.50	50.70	Sol
10:31	3	8	33.60	50.50	Sol
10:32	2	8	34.20	50.90	Sol
10:33	3	8	33.40	50.40	Sol
10:34	2	8	32.80	50.50	Sol
10:35	3	8	34.50	51.00	Sol
10:36	3	8	34.30	50.10	Sol
10:37	2	8	34.50	49.30	Sol
10:38	20	0	34.00	49.60	Sombra
10:39	20	0	34.60	49.70	Sombra
10:40	19	0	33.90	51.30	Sombra
10:41	19	0	33.40	53.30	Sombra

3.2 Modelos de Regressão Logística Multinomial

O modelo de regressão logística multinomial foi adotado para modelar o problema descrito na seção (3.1) devido à natureza da variável resposta. As variáveis do conjunto X podem ser escritas como soma linear de seus efeitos. A variável Y é de natureza multinomial adotando três categorias. Quinze modelos foram construídos envolvendo a combinação das quatro variáveis explicativas, cuja estrutura é descrita na tabela (3.3).

Tabela 3.3: Modelos

Modelo	**Descrição**	**Estrutura Linear**
M_1	Modelo de uma variável: Luz	$Y \backsim X_1$
M_{12}	Modelo com duas variáveis: Luz e UV	$Y \backsim X_1 + X_2$
M_{123}	Modelo com três variáveis: Luz, UV e Temperatura	$Y \backsim X_1 + X_2 + X_3$
M_{1234}	Modelo Saturado com quatro variáveis	$Y \backsim X_1 + X_2 + X_3 + X_4$
M_{124}	Modelo com três variáveis: Luz, UV e Umidade	$Y \backsim X_1 + X_2 + X_4$
M_{13}	Modelo com duas variáveis: Luz e Temperatura	$Y \backsim X_1 + X_3$
M_{134}	Modelo com três variáveis: Luz, Temperatura e Umidade	$Y \backsim X_1 + X_3 + X_4$
M_{14}	Modelo com duas variáveis: Luz e Umidade	$Y \backsim X_1 + X_4$
M_2	Modelo com uma variável: UV	$Y \backsim X_2$
M_{23}	Modelo com duas variáveis: UV e Temperatura	$Y \backsim X_2 + X_3$
M_{234}	Modelo com três variáveis: UV, Temperatura e Umidade	$Y \backsim X_2 + X_3 + X_4$
M_{24}	Modelo com duas variáveis: UV e Umidade	$Y \backsim X_2 + X_4$
M_3	Modelo com uma variável: Temperatura	$Y \backsim X_3$
M_{34}	Modelo com duas variáveis: Temperatura e Umidade	$Y \backsim X_3 + X_4$
M_4	Modelo com uma variável: Umidade	$Y \backsim X_4$
M_0	Modelo sem variáveis com interceptos	$Y \backsim 1$

Cada modelo assume a distribuição multinomial da variável Y bem como o uso da função *logit* como função de ligação definida pela expressão (2.29). O método de estimação de parâmetros é o da máxima verossimilhança descrita na proposição (2.11). O método numérico para resolução das equações de verossimilhança é descrito em (2.42).

O teste da razão de verossimilhança e de Wald foram aplicados para cada modelo. A razão de verossimilhança testada foi baseada na estatística G definida em (2.45), comparando o modelo selecionado com o modelo M_0 sem variáveis preditoras. Os valores das estatísticas pseudo-R^2 foram obtidas através das expressões (2.71), (2.72) e (2.73).

Para cada modelo foi construída uma tabela de classificação. A regra para classificar uma observação em uma classe que assuma o valor de Sol, Sombra ou Nublado, é tomar a maior das três probabilidades da observação no modelo.

Os resumos para cada modelo podem ser encontrados no apêndice desse trabalho. Em cada

tabela "LogLik"indica o valor maximizado do logaritimo de máxima verossimilhança, $\mathcal{X}^2$ indica o valor da estatística G que tem distribuição qui-quadrado, "Df"indica os graus de liberdade dessa distribuição. Um p-valor pequeno o suficiente implica na rejeição da hipótese nula. Entre esses quinze modelos é possível perceber diferenças na significância de seus parâmetros bem como seu poder de previsão. No entanto é necessário o estabelecimento de critérios que definam o modelo que melhor se ajuste ao problema. O **critério de informação de Akaike** (AIC) e o **critério de informação Bayesiano** (BIC) permitem a comparação entre modelos [14].

O Critério de Informação de Akaike é definido como:

$$AIC = -2\log(L(\hat{\theta})) + 2k \tag{3.1}$$

onde $L(\hat{\theta})$ é o valor máximizado da função de verossimilhança e k é o número de parâmetros a ser estimado.

O Critério de Informação Bayesiano é definido como:

$$BIC = -2\log(L(\hat{\theta})) + k\log(n) \tag{3.2}$$

onde $L(\hat{\theta})$ é o valor máximizado da função de verossimilhança, k o número de parâmetros a ser estimado e n o tamanho da amostra.

Enquanto o AIC penaliza um modelo por ter muitos parâmetos, o BIC penaliza ainda mais por incluir o tamanho da amostra. O modelo mais desejável é o que apresenta o menor AIC ou BIC em relação aos demais modelos. Entre os quinze modelos o que apresenta o menor BIC é o modelo M_{234} e o de menor AIC é o modelo M_{1234}. A tabela (3.4) apresenta os valores dos critérios AIC e BIC para cada modelo.

Tabela 3.4: Valores de AIC e BIC dos modelos

Modelo	**AIC**	**BIC**	**Modelo**	**AIC**	**BIC**
M_1	1001.2903	1018.3286	M_2	320.6873	337.7256
M_{12}	323.5778	349.1353	M_{23}	278.8389	304.3964
M_{123}	254.2510	288.3276	M_{234}	199.5324	**233.6091**
M_{1234}	**196.3966**	238.9924	M_{24}	241.4292	266.9867
M_{124}	241.3894	275.4661	M_3	1016.7660	1033.8043
M_{13}	966.7576	992.3151	M_{34}	1014.9097	1040.4672
M_{134}	963.6382	997.7149	M_4	1047.2396	1064.2779
M_{14}	977.7469	1003.3044			

3.3 Resultados

Entre os quinze modelos dois modelos se destacaram em relação à significância de variáveis, poder de previsão e ajuste ao conjunto de dados. O modelo saturado com todas as variáveis, M_{1234}, apresentou o menor AIC e o modelo M_{234} o menor BIC.

3.3.1 Modelo Saturado M_{1234}

O modelo M_{1234} faz uso de todas as variáveis explicativas e possui variável dependente Y assumindo valores dentro do conjunto $\{Sol, Sombra, Nublado\}$. O resumo de todas as informações desse modelo estão descritas na tabela (3.5).

O teste da razão de verossimilhança resultou significativo (p-valor= $2.2e^{-16}$) indicando que o modelo estimado pode ser útil na discriminação dos três estados de nível de sombreamento. Os valores *Pseudo-*R^2 tem uma interpretação complexa, porém seguem uma regra básica: quanto maior o valor, melhor o ajuste do modelo. Dentre as três medidas, é notável o valor de *Nagelkerke*, medida no intervalo [0;1], que apresentou um valor muito próximo do máximo (0.945).

As estimativas dos parâmetros do modelo estão contidas na tabela (3.5). Dez parâmetros foram estimados devido nesse caso haver três classes e quatro variáveis. As duas funções *logit* estimadas foram:

$$\begin{aligned} g_1(X) &= 48.7078 + 0.0106X_1 - 29.6229X_2 - 0.3337X_3 - 0.4336X_4 \quad \text{(Classe Sombra)} \\ g_2(X) &= 50.9773 - 0.0526X_1 - 2.0973X_2 - 0.6863X_3 - 0.3048X_4 \quad \text{(Classe Nublado)} \end{aligned} \tag{3.3}$$

A classe Sol foi tomada como referência e, portanto $g_3(X) = 0$. De acordo com o teste de Wald, 4 parâmetros não resultaram significativos com $p - valor > 0.05$, entretanto verifica-se que com exceção da variável luz, todas as demais variáveis têm coeficientes significativos em pelo menos uma das equações. Retirar a variável X_1 do modelo resulta na obtenção do modelo M_{234}, modelo que apresenta o menor BIC. A aplicação do modelo consiste em inserir valores de x nas funções *logit* para obtenção das probabilidades de pertencer as classes:

$$\begin{aligned} P(Y = Sombra|\boldsymbol{x}) &= \frac{exp\{g_1(x)\}}{1 + exp\{g_1(x)\} + exp\{g_2(x)\}} \\ P(Y = Nublado|\boldsymbol{x}) &= \frac{exp\{g_2(x)\}}{1 + exp\{g_1(x)\} + exp\{g_2(x)\}} \\ P(Y = Sol|\boldsymbol{x}) &= 1 - P(Y = Sombra|\boldsymbol{x}) - P(Y = Nublado|\boldsymbol{x}) \end{aligned} \tag{3.4}$$

Tabela 3.5: Resumo do Modelo M_{1234}

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_{1234}	-88.2	912.2	8	2.2e-16

Pseudo R Quadrado	
Cox e Snell	0.836
Nagelkerke	0.945
McFadden	0.838

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	48.7078	18.2965	2.6621	0.0078	[12847,84568]
	X_1	0.0106	0.0557	0.1908	0.8487	[-0.099,0.120]
	X_2	-29.6229	391.1842	-0.0757	0.9396	[-796330,737084]
	X_3	-0.3337	0.3810	-0.8758	0.3811	[-1080,0.413]
	X_4	-0.4336	0.1032	-4.2000	0.0000	[-0.636,-0.231]
Nublado	Intercepto	50.9773	8.0204	6.3560	0.0000	[35258,66697]
	X_1	-0.0526	0.0361	-1.4588	0.1446	[-0.123,0.018]
	X_2	-2.0973	0.2706	-7.7492	0.0000	[-2628,-1567]
	X_3	-0.6963	0.1338	-5.2048	0.0000	[-0.959,-0.434]
	X_4	-0.3048	0.0504	-6.0427	0.0000	[-0.404,-0.206]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	233	0	16	93,57%
Sombra	0	103	2	98,10%
Nublado	11	2	156	92,31%
% Geral	46,65%	20,08%	33,27%	**94,07%**

Como exemplo, uma amostra $\boldsymbol{x}$ com valores de luminosidade de 4, a medida de radiação UV de 5, uma temperatura de 28.0° e umidade relativa do ar medida em 81.8% teria as seguintes probabilidades de classificação:

$$\begin{aligned} P(Y = Sombra|\boldsymbol{x}) &\approx 0 \\ P(Y = Nublado|\boldsymbol{x}) &\approx 0.0156 \\ P(Y = Sol|\boldsymbol{x}) &\approx 0.9844 \end{aligned} \tag{3.5}$$

Portanto uma leitura com tais medições seria classificado como Sol. A tabela (3.5) apresenta a tabela de classificação utilizando todas as 505 observações, onde percebe-se uma habilidade satisfatória do modelo para classificação com taxa de acerto de 94,07%.

3.3.2 Modelo M_{234}

O modelo M_{234} apresentou o menor valor de BIC e a não significância estatística da variável X_1 no modelo M_{1234} motiva a análise desse modelo ajustado. M_{234} faz uso das variáveis que envolvem a radiação UV, a temperatura e a umidade relativa do ar. O resumo de todas as informações desse modelo estão descritas na tabela (3.6).

O teste da razão de verossimilhança resultou significativo (p-valor= $2.2e^{-16}$) indicando que o modelo é útil para a discriminação do nível de sombreamento. O valor do coeficiente de determinação de *Nagelkerke* foi próximo do valor máximo, 0.943. Seu poder de previsão foi avaliado em 94,26%.

As estimativas dos oito parâmetros do modelo são apresentadas na tabela (3.6). As duas funções *logit* estimadas foram:

$$\begin{aligned} g_1(X) &= 50.9690 - 29.5118X_2 - 0.3647X_3 - 0.42999X_4 \quad \text{(Classe Sombra)} \\ g_2(X) &= 44.8914 - 2.1470X_2 - 0.5726X_3 - 0.2673X_4 \quad \text{(Classe Nublado)} \end{aligned} \tag{3.6}$$

A classe Sol foi tomada como referência e, portanto $g_3(x) = 0$. De acordo com o teste de Wald apenas dois parâmetros estimados não resultaram significativos com p-valor > 0.05, entretanto verifica-se que todas as variáveis têm coeficientes significativos em uma das duas das equações, o que constitui uma situação desejável. O teste de razão de verossimilhança (TRV) para as variáveis X_2 e X_3 permitem confirmar a significância das variáveis conforme as tabelas (3.7) e (3.8).

A aplicação do modelo ocorre de forma similar ao modelo saturado, onde consiste em inserir valores de x nas funções *logit* para obtenção das probabilidades de pertencer as classes:

$$\begin{aligned} P(Y = Sombra|\boldsymbol{x}) &= \frac{exp\{g_1(x)\}}{1 + exp\{g_1(x)\} + exp\{g_2(x)\}} \\ P(Y = Nublado|\boldsymbol{x}) &= \frac{exp\{g_2(x)\}}{1 + exp\{g_1(x)\} + exp\{g_2(x)\}} \\ P(Y = Sol|\boldsymbol{x}) &= 1 - P(Y = Sombra|\boldsymbol{x}) - P(Y = Nublado|\boldsymbol{x}) \end{aligned} \tag{3.7}$$

Como exemplo, uma amostra $\boldsymbol{x}$ com valores de luminosidade de 4, a medida de radiação UV de 5, uma temperatura de 28.0° e umidade relativa do ar medida em 81.8% teria as seguintes probabilidades de classificação:

$$\begin{aligned} P(Y = Sombra|\boldsymbol{x}) &\approx 0 \\ P(Y = Nublado|\boldsymbol{x}) &\approx 0.0231 \\ P(Y = Sol|\boldsymbol{x}) &\approx 0.9769 \end{aligned} \tag{3.8}$$

Tabela 3.6: Resumo do Modelo M_{234}

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_{234}	-91.77	905.06	6	2.2e-16

Pseudo R Quadrado	
Cox e Snell	0.833
Nagelkerke	0.943
McFadden	0.831

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	50.9690	11.6419	4.3781	0.0000	[28.151,73.787]
	X_2	-29.5118	401.3081	-0.0735	0.9414	[-816.061,757.038]
	X_3	-0.3647	0.2463	-1.4807	0.1387	[-0.848,0.118]
	X_4	-0.4299	0.0686	-6.2648	0.0000	[-0.564,-0.295]
Nublado	Intercepto	44.8914	6.7896	6.6118	0.0000	[31.584,58.199]
	X_2	-2.1470	0.2833	-7.5774	0.0000	[-2.702,-1.592]
	X_3	-0.5726	0.1020	-5.6136	0.0000	[-0.773,-0.373]
	X_4	-0.2673	0.0435	-6.1413	0.0000	[-0.353,-0.182]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	233	0	16	93,57%
Sombra	0	101	4	96,19%
Nublado	9	1	159	94,08%
% Geral	46,27%	19,50%	34,23%	**94,26%**

Tabela 3.7: TRV do modelo M_{234} em relação à variável X_2

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_{234}	-91.77			
M_{34}	-501.45	819.38	2	2.2e-16

Tabela 3.8: TRV do modelo M_{234} em relação à variável X_3

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_{234}	-91.77			
M_{24}	-114.715	45.897	2	1.081e-10

Nesse modelo estimado, uma leitura com tais medições seria classificado como Sol, que corresponde a realidade. A tabela (3.6) apresenta a tabela de classificação utilizando todas as 505 observações, onde percebe-se uma habilidade satisfatória do modelo para classificação com taxa de acerto de 94,26%.

O valor de BIC do modelo M_{234} juntamente com a não significância estatística da variável X_1, que representa a luminosidade ambiente, leva a concluir que M_{234} tem o melhor desempenho quando comparado com os quinze modelos estudados. Enquanto o modelo M_{1234} possui uma taxa de acerto de 94,07% com todas as variáveis explicativas, existe um aumento da taxa de acerto no modelo M_{234} com a retirada de uma variável. Na análise estatística um modelo com menor quantidade de parâmetros é desejável pois reduz a possibilidade de erros durante o processo de estimação. Além disso, uma quantidade menor de variáveis explanatórias iniciais representa uma redução na quantidade de sensores de captação dessas variáveis e, como consequência, uma redução no custo da plataforma de coleta de dados.

Capítulo 4

Conclusão

A regressão logística multinomial demonstrou grande valor na análise de variáveis nominais, apresentando características interessantes para a modelagem de diversos problemas. Uma das características, de especial importância nesse trabalho foi a possibilidade de estimar a probabilidade de uma observação pertencer a uma determinada classe. Isso resulta que a aplicação de um modelo em um novo conjunto de dados seja simples, pois, após a estimação de parâmetros, envolve inserir as variáveis da observação em uma equação.

De acordo com os modelos estudados, percebeu-se a não significância estatística da variável ambiental Luminosidade para indicar o conforto térmico de bovinos pelo nível de sombreamento. Tal variável ambiental foi representada nos modelos pela variável aleatória X_1. Não incluir essa variável no modelo resultou no modelo M_{234} que faz uso de todas as demais variáveis ambientais. A exclusão dessa variável representa a retirada de um dos sensores da plataforma de captação de variáveis e redução nos custos de produção da plataforma.

O modelo M_{234} apresentou um melhor desempenho em relação ao modelo saturado M_{1234}. Com uma variável a menos foi obtido um modelo com melhor capacidade de previsão. De modo geral, os modelos apresentaram uma boa taxa de classificação e significância estatística dos estimadores que contribui em seu poder de previsão.

A aplicação do modelo M_{234} em novos conjuntos de dados resumiu-se em inserir os valores de cada observação e calcular a probabilidade afim de obter o nível de sombreamento mais provável.

Como trabalhos futuros pode-se citar a análise da estrutura linear entre as variáveis independentes que resulte em um modelo que, ainda não sendo um modelo linear generalizado, possa modelar melhor a situação.

Referências Bibliográficas

[1] *MEMÓRIA, J.M.P.* Breve História da Estatística. Embrapa Informação Tecnológica: Brasília, DF, 2004

[2] *STEWART I.* 17 Equações que Mudaram o Mundo. Rio de Janeiro: Zahar, 2013.

[3] *CRAMER J.S.* The Origins of Logistic Regression. Cambridge University Press, 2002

[4] *BRASIL. Secretaria de Educação Fundamental* Parâmetros curriculares nacionais : matemática / Secretaria de Educação Fundamental. Brasília : MEC/SEF, 1997.

[5] *BRASIL.* PCN+ Ensino Médio. Brasília: Ministério da Educação,2002.

[6] *BRASIL.* Lei de Diretrizes e Bases da Educação Nacional, Lei 9.394, de 20 de dezembro de 1996 (LDB).

[7] *CAMPO GRANDE. Secretaria Municipal de Educação.* Referencial Curricular da Rede Municipal de Ensino. Campo Grande, MS, 2008

[8] *MATO GROSSO DO SUL. Secretaria de Estado de Educação.* Referencial Curricular da Rede Estadual de Ensino de Mato Grosso do Sul Fundamental. Campo Grande, MS, 2012

[9] *MATO GROSSO DO SUL. Secretaria de Estado de Educação.* Referencial Curricular da Rede Estadual de Ensino de Mato Grosso do Sul Ensino Médio. Campo Grande, MS, 2012

[10] *BRASIL. Ministério da Educação. Secretaria de Educação Básica.* Orientações curriculares para o ensino médio, vol. 2: Ciências da natureza, matemática e suas tecnologias. Brasília: MEC. 2006

[11] *AMARAL, M.R.S.; CESÁRIO, C.V.; GONÇALVES, G.A.; SCHULTZ, L.T.; CUNHA, K.P.V.;SOUZA, T.S.;* Apostila do curso de extensão: Software Estatístico Livre R.IME/UERJ: Rio de Janeiro, RJ, 2010

[12] *R Developement Core Team (2016)*; R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

[13] *SARAIVA, E.F.* Conceitos básicos de probabilidade e estatística. Notas de aulas,INMA/UFMS: Campo Grande, MS, 2015

[14] *AGRESTI, A.* Categorical Data Analysis. John Wiley, New York, 1990.

[15] *CORDEIRO, G. M e DEMÉTRIO, C.G.B* Modelos Lineares Generalizados e Extensões, UFRPE-USP, 2010

[16] *HOSMER, D.W. e LEMESHOW, S.* Applied Logistic Regression. John Wiley, New York, 2000

[17] *FIGUEIRA,C.V.* Modelos de Regressão Logística, 2006, Dissertação (Mestrado em Matemática) - UFRS, 2006.

[18] *COX D.R. e SNELL, E.J.* Analysis of Binary Data. Chapman & Hall, Londres, 1989

[19] *NAGELKERKE, N.J.D.* "A Note on a general definition of the coefficient of determination". Biometrika, Vol 78, 1991

[20] *Alves, F. V., de Almeida, R. G., Laura, V. A., de Oliveira, C. C.* Ambiente y bienestar de bovinos de carne en sistemas integrados Cultivos-Ganadería-Florestal en Brasil, II Congreso Colombiano y 1 Seminario Internacional Silvopastoreo, 2012.

[21] *SILVA, R.G.* Introdução à bioclimatologia animal. São Paulo: Nobel, 2000.

[22] *GLASER, F.D.* Aspectos comportamentais de bovinos das raças Angus, Caracu e Nelore a pasto frente à disponibilidade de recursos de sombra e água para imersão. 2008. 117 f. Tese (Doutorado em Zootecnia), Universidade de São Paulo (USP), Pirassununga, 2008.

[23] *MARIN,C.* Sistema para Aquisição e Processamento de Temperatura Cutânea em Bovinos e Variáveis Ambientais, 2016, Dissertação - UFMS, 2016.

[24] *Thomas W. Yee (2017)* VGAM: Vector Generalized Linear and Additive Models. R package version 1.0-3. URL https://CRAN.R-project.org/package=VGAM

Apêndice A

Resumo dos Modelos

Este apêndice contém os resumos de cada um dos quinze modelos construídos descritos no capítulo 3. Cada resumo detalha as estimativas dos parâmetros, testes de significância e tabela de classificação para cada modelo.

Tabela A.1: Resumo do Modelo M_1

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_1	-496.65	95.303	2	2.2e-16

Pseudo R Quadrado	
Cox e Snell	0.172
Nagelkerke	0.195
McFadden	0.088

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	-1.8149	0.1788	-10.1499	2e-16	[-2,174,-1,473]
	X_1	0.1103	0.1376	-5.0582	4.23e-07	[-0.966,-0.426]
Nublado	Intercepto	-0.6960	0.0174	6.3543	2.09e-10	[0.076,0.144]
	X_1	0.0546	0.0170	3.2114	0.0013	[0.021,0.088]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	242	7	0	97,19%
Sombra	47	45	13	42,86%
Nublado	148	10	11	6,51%
% Geral	83,56%	11,85%	4,59%	**56,98%**

Tabela A.2: Resumo do modelo M_{12}

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_{12}	-155.79	777.02	4	2.2e-16

Pseudo R Quadrado	
Cox e Snell	0.785
Nagelkerke	0.888
McFadden	0.714

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	8.5674	0.9286	9.2258	2e-16	[6.747,10.387]
	X_1	0.0362	0.0306	1.1856	0.2358	[-0.024,0.096]
	X_2	-19.9845	530.7110	-0.0377	0.9700	[-1060.1 1020.1]
Nublado	Intercepto	7.2054	0.8580	8.3982	2e-16	[5.524,8.887]
	X_1	0.0265	0.0273	0.9706	0.3317	[-0.027,0.080]
	X_2	-1.3457	0.1495	-9.0037	2e-16	[-1.639,-1.053]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	223	0	26	89,56%
Sombra	0	105	0	100,00%
Nublado	29	22	118	69,82%
% Geral	48,18%	24,28%	27,53%	**85,28%**

Tabela A.3: Resumo do Modelo M_{123}

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_{123}	-119.13	850.34	6	2.2e-16

Pseudo R Quadrado	
Cox e Snell	0.814
Nagelkerke	0.921
McFadden	0.781

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	-18.5600	5.1314	-3.617	0.000298	[-28.618,-8.503]
	X_1	0.1252	0.0421	2.970	0.002983	[0.043,0.208]
	X_2	-31.4042	403.5687	-0.078	0.937974	[-822.384 759.576]
	X_3	0.8312	0.1602	5.186	2.14e-07	[0.517,1.145]
Nublado	Intercepto	7.4756	1.4445	5.175	2.28e-07	[4.644,10.307]
	X_1	0.0252	0.0279	0.905	0.365209	[-0.029,0.080]
	X_2	-1.3399	0.1513	-8.854	2e-16	[-1.636,-1.043]
	X_3	-0.0084	0.0360	-0.235	0.813953	[-0.079,0.062]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	223	0	26	89,56%
Sombra	0	102	3	97,14%
Nublado	29	3	137	81,07%
% Geral	48,18%	20,08%	31,74%	**88,34%**

Tabela A.4: Resumo do Modelo M_{124}

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_{124}	-112.69	863.2	6	2.2e-16

Pseudo R Quadrado	
Cox e Snell	0.819
Nagelkerke	0.926
McFadden	0.793

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	25.5904	3.2623	7.8442	0.0000	[19.196,31.984]
	X_1	0.1016	0.0552	1.8397	0.0658	[-0.007,0.210]
	X_2	-27.3685	445.9395	-0.0614	0.9511	[-901.394,846.657]
	X_4	-0.2698	0.0449	-6.0098	0.0000	[-0.358,-0.182]
Nublado	Intercepto	11.1169	1.6903	6.5770	0.0000	[7.804,14.430]
	X_1	0.0327	0.0278	1.1784	0.2386	[-0.022,0.087]
	X_2	-1.5836	0.1873	-8.4555	0.0000	[-1.951,-1.217]
	X_4	-0.0528	0.0170	-3.0981	0.0019	[-0.086,-0.019]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	224	0	25	89,96%
Sombra	0	103	2	98,10%
Nublado	32	1	136	80,47%
% Geral	48,95%	19,89%	31,17%	**88,53%**

Tabela A.5: Resumo do Modelo M_{13}

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_{13}	-477.38	133.84	4	2.2e-16

Pseudo R Quadrado	
Cox e Snell	0.233
Nagelkerke	0.263
McFadden	0.123

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	1.2432	0.9538	1.3034	0.1924	[-0.626,3.113]
	X_1	0.0789	0.0178	4.4475	0.0000	[0.044,0.114]
	X_3	-0.0762	0.0244	-3.1259	0.0018	[-0.124,-0.028]
Nublado	Intercepto	3.6481	0.7628	4.7826	0.0000	[2.153,5.143]
	X_1	0.0144	0.0177	0.8168	0.4140	[-0.020,0.049]
	X_3	-0.1106	0.0195	-5.6783	0.0000	[-0.149,-0.072]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	175	7	67	70,28%
Sombra	39	48	18	45,71%
Nublado	121	3	45	26,63%
% Geral	64,05%	11,09%	24,86%	**51,24%**

Tabela A.6: Resumo do Modelo M_{134}

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_{134}	-473.82	140.96	6	2.2e-16

Pseudo R Quadrado	
Cox e Snell	0.244
Nagelkerke	0.275
McFadden	0.129

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	-1.8129	3.0474	-0.5949	0.5519	[-7.786,4.160]
	X_1	0.0859	0.0178	4.8314	0.0000	[0.051,0.121]
	X_3	-0.0189	0.0572	-0.3305	0.7410	[-0.131,0.093]
	X_4	0.0189	0.0206	0.9204	0.3574	[-0.021,0.059]
Nublado	Intercepto	8.2941	2.6316	3.1517	0.0016	[3.136,13.452]
	X_1	0.0112	0.0174	0.6452	0.5188	[-0.023,0.045]
	X_3	-0.1959	0.0500	-3.9157	0.0001	[-0.294,-0.098]
	X_4	-0.0325	0.0181	-1.7996	0.0719	[-0.068,0.003]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	164	7	78	65,86%
Sombra	43	50	12	47,62%
Nublado	97	14	58	34,32%
% Geral	58,13%	13,58%	28,30%	**52,01%**

Tabela A.7: Resumo do Modelo M_{14}

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_{14}	-482.87	122.85	4	2.2e-16

Pseudo R Quadrado	
Cox e Snell	0.216
Nagelkerke	0.244
McFadden	0.113

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	-3.0616	0.4152	-7.3736	0.0000	[-3.875,-2.248]
	X_1	0.0818	0.0167	4.9070	0.0000	[-2.652,-1.397]
	X_4	0.0316	0.0086	3.6667	0.0002	[0.049,0.114]
Nublado	Intercepto	-2.0243	0.3202	-6.3221	0.0000	[-0.010,0.057]
	X_1	0.0237	0.0171	1.3843	0.1663	[0.015,0.049]
	X_4	0.0337	0.0070	4.7779	0.0000	[0.020,0.047]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	201	7	41	80,72%
Sombra	46	44	15	41,90%
Nublado	125	10	34	20,12%
% Geral	71,13%	11,66%	17,21%	**53,35%**

Tabela A.8: Resumo do Modelo M_2

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_2	-156.34	775.91	2	2.2e-16

Pseudo R Quadrado	
Cox e Snell	0.785
Nagelkerke	0.888
McFadden	0.713

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	8.8415	0.8927	99046	0.0000	[7.092,10.591]
	X_2	-20.1242	533.2561	-0.0377	0.9699	[-1065.287,1025.039]
Nublado	Intercepto	7.2778	0.8619	84439	0.0000	[5.589,8.967]
	X_2	-1.3416	0.1485	-90322	0.0000	[-1.633,-1.050]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	223	0	26	89,56%
Sombra	0	105	0	100,00%
Nublado	29	22	118	69,82%
% Geral	48,18%	24,28%	27,53%	**85,28%**

Tabela A.9: Resumo do Modelo M_{23}

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_{23}	-133.42	821.75	4	2.2e-16

Pseudo R Quadrado	
Cox e Snell	0.804
Nagelkerke	0.909
McFadden	0.755

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	-6.0817	3.3049	-1.8402	0.0657	[-12.559,0.396]
	X_2	-26.7840	416.3659	-0.0643	0.9487	[-842.846,789.278]
	X_3	0.4995	0.1111	4.4951	0.0000	[0.282,0.717]
Nublado	Intercepto	7.6392	1.4247	5.3621	0.0000	[4.847,10.431]
	X_2	-1.3335	0.1504	-8.8666	0.0000	[-1.628,-1.039]
	X_3	-0.0115	0.0357	-0.3236	0.7463	[-0.081,0.058]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	223	0	26	89,56%
Sombra	0	96	9	91,43%
Nublado	29	9	131	77,51%
% Geral	48,18%	20,08%	31,74%	**86,04%**

Tabela A.10: Resumo do Modelo M_{24}

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_{24}	-114.71	859.16	4	2.2e-16

Pseudo R Quadrado	
Cox e Snell	0.818
Nagelkerke	0.925
McFadden	0.789

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	26.1662	3.3199	7.8815	0.0000	[19.659,32.673]
	X_2	-27.6705	445.7192	-0.0621	0.9505	[-901.264,845.923]
	X_4	-0.2555	0.0431	-5.9292	0.0000	[-0.340,-0.171]
Nublado	Intercepto	11.1482	1.6911	6.5923	0.0000	[7.834,14.463]
	X_2	-1.5760	0.1859	-8.4762	0.0000	[-1.940,-1.212]
	X_4	-0.0519	0.0169	-3.0749	0.0021	[-0.085,-0.019]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	224	0	25	89,96%
Sombra	0	101	4	96,19%
Nublado	32	2	135	79,88%
% Geral	48,95%	19,69%	31,36%	**87,95%**

Tabela A.11: Resumo do Modelo M_3

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_3	-504.38	79.827	2	2.2e-16

Pseudo R Quadrado	
Cox e Snell	0.146
Nagelkerke	0.165
McFadden	0.073

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	4.9597	0.8101	6.1223	0	[3.372,6.547]
	X_3	-0.1603	0.0228	-7.0258	0	[-0.205,-0.116]
Nublado	Intercepto	3.6096	0.6482	5.5685	0	[2.339,4.880]
	X_3	-0.1073	0.0174	-6.1596	0	[-0.141,-0.073]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	169	1	79	67,87%
Sombra	61	2	42	1,90%
Nublado	121	5	43	25,44%
% Geral	67,11%	1,53%	31,36%	**40,92%**

Tabela A.12: Resumo do Modelo M_{34}

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_{34}	-501.45	85.683	4	2.2e-16

Pseudo R Quadrado	
Cox e Snell	0.156
Nagelkerke	0.177
McFadden	0.079

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	10.0608	2.3850	4.2184	0.0000	[5.386,14.735]
	X_3	-0.2515	0.0466	-5.3955	0.0000	[-0.343,-0.160]
	X_4	-0.0387	0.0167	-2.3161	0.0206	[-0.071,-0.006]
Nublado	Intercepto	5.5763	2.1571	2.5851	0.0097	[1.348,9.804]
	X_3	-0.1421	0.0406	-3.5003	0.0005	[-0.222,-0.063]
	X_4	-0.0149	0.0155	-0.9658	0.3341	[-0.045,0.015]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	169	7	73	67,87%
Sombra	59	10	36	9,52%
Nublado	110	0	59	34,91%
% Geral	64,63%	3,25%	32,12%	**45,51%**

Tabela A.13: Resumo do Modelo M_4

Informação de ajuste de modelo				
Modelo	**LogLik**	$\mathcal{X}^2$	**Df**	**p-Valor**
M_0	-544.30			
M_1	-519.62	49.354	2	1.919e-11

Pseudo R Quadrado	
Cox e Snell	0.093
Nagelkerke	0.105
McFadden	0.045

Estimativa de Parâmetros - Categoria de Referência=Sol

Classe		**Estimativa**	**SE**	**Wald**	**p-valor**	**IC 95%**
Sombra	Intercepto	-2.8840	0.3866	-7.4599	0	[-3.642,-2.126]
	X_1	0.0444	0.0078	5.7163	0	[0.029,0.060]
Nublado	Intercepto	-2.8840	0.3866	-7.4599	0	[-2.683,-1.421]
	X_1	0.0373	0.0067	5.5277	0	[0.024,0.051]

Tabela de Classificação				
	Predito			
Observado	Sol	Sombra	Nublado	% Correto
Sol	192	0	57	77,11%
Sombra	71	0	34	0,00%
Nublado	124	0	45	26,63%
% Geral	74,00%	0,00%	26,00%	**45,32%**