

## PREDIÇÃO DO NÍVEL DE SOMBREAMENTO UTILIZANDO UM MODELO DE REGRESSÃO LOGÍSTICA MULTINOMIAL EM SISTEMAS DE CRIAÇÃO DE BOVINOS DE CORTE

Rubia Mara de Oliveira SANTOS<sup>1</sup>  
Erlandson Ferreira SARAIVA<sup>1</sup>  
Thiago Farias ARCE<sup>1</sup>  
Ricardo Ribeiro dos SANTOS<sup>2</sup>

- RESUMO: Nos últimos anos, as consequências do estresse térmico em bovinos de corte têm sido assunto de interesse para pesquisadores, investidores e criadores. Isto se deve ao fato do estresse calórico impactar na diminuição do ganho de peso dos animais e aumento do tempo de abate e do custo da produção. Dessa forma, ocasionando perdas econômicas aos criadores/investidores. Neste artigo, apresentamos uma plataforma eletrônica desenvolvida para captar medidas ambientais em sistemas de criação de gado de corte. A partir dos dados captados pela plataforma eletrônica ajustamos quinze modelos de regressão logística multinomial para predição do nível de sombreamento (Sol, Nublado e Sombra) que um bovino se encontra. Para selecionar o melhor modelo dentre os ajustados, consideramos os critérios de seleção de modelos AIC e BIC. Obtemos, como resultado, um modelo com alta capacidade de predição correta. Do ponto de vista prático/econômico a plataforma eletrônica juntamente com o modelo ajustado se mostram promissores na identificação de animais que estão à procura de sombra, sendo este um indicativo de estresse térmico.
- PALAVRAS-CHAVE: Estresse térmico; regressão logística multinomial; seleção de modelos; predição.

---

<sup>1</sup>Universidade Federal de Mato Grosso do Sul - UFMS, Instituto de Matemática, CEP: 79070-900, Campo Grande, MS, Brasil. E-mail: [rubia.oliveira@ufms.br](mailto:rubia.oliveira@ufms.br); [erlandson.saraiva@ufms.br](mailto:erlandson.saraiva@ufms.br); [thiagocarce@gmail.com](mailto:thiagocarce@gmail.com).

<sup>2</sup>Universidade Federal de Mato Grosso do Sul - UFMS, Faculdade de Computação, CEP: 79070-900, Campo Grande, MS, Brasil. E-mail: [ricardo@facom.ufms.br](mailto:ricardo@facom.ufms.br)

## 1 Introdução

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 (ALVES *et al.*, 2012). De acordo com Mitlöhner (2002), isto se deve ao fato de que um animal em estresse térmico reduz a sua produção de calor metabólico reduzindo a ingestão de alimentos; com isso acarretando uma diminuição do ganho de peso, aumentando o tempo de abate e o custo de produção.

Segundo Glaser (2008) uma alternativa para evitar o estresse térmico é o uso do sombreamento que diminui a incidência da radiação solar sobre o animal. No entanto, a umidade do ar e a temperatura, de acordo com Silva (2000), são os maiores responsáveis pelo conforto térmico animal.

Além das variáveis temperatura e umidade do ar, a luminosidade e a radiação UV também são importantes para avaliar o nível de sombreamento do bovino e identificar uma possível situação de estresse térmico (MARIN, 2016). Como o estresse térmico apresenta grande impacto na produção, há um crescente interesse em métodos e técnicas estatísticas/computacionais que permitam identificar quando um animal está sob estresse térmico. Dessa forma, possibilitar uma intervenção de profissionais especializados de modo a proporcionar uma melhor qualidade de vida ao animal e conseqüentemente diminuir as perdas financeiras.

Neste artigo, apresentamos um método automatizado para predição do nível de sombreamento que o animal se encontra a partir de dados de variáveis ambientais. Esta abordagem é motivada pelo fato de que um animal em estresse térmico tende a procurar sombra. Para isto, foi desenvolvido pelo Laboratório de Sistemas Computacionais de Alto Desempenho (LSCAD) da Faculdade de Computação (FACOM) da Universidade Federal de Mato Grosso do Sul (UFMS) uma plataforma eletrônica para coletar e analisar medidas referentes as variáveis ambientais: temperatura, luminosidade do ambiente, umidade relativa do ar e radiação Ultra-Violeta.

A solução de hardware e software desenvolvida pelo LSCAD tem por objetivo identificar as variações de temperatura na pele do animal, bem como a indicação da sua busca por sombra. Nosso objetivo é o desenvolvimento de um modelo estatístico para prever o nível de sombreamento (sol, nublado e sombra) incidente sobre o animal a partir das medidas ambientais captadas pela plataforma. De posse desta informação, o profissional especializado da área (veterinário e/ou zootecnista) avaliará esta informação juntamente com a temperatura da pele do animal para decidir pela intervenção ou não com o objetivo de minimizar os efeitos do estresse térmico.

Para predição do nível de sombreamento, consideramos o ajuste de um modelo de regressão logística multinomial. Detalhes sobre o ajuste de um modelo de regressão logística multinomial (RLM) podem ser encontrados em Agresti (2012). Além disso, na literatura podemos encontrar diversas aplicações deste modelo nas mais diversas áreas do conhecimento. Entre eles, von Borries *et al.*, (2018) utilizam o modelo RLM em conjunto com análise de componentes principais em uma análise

de avaliação sensorial; Zhu e Hastie (2004) propõem um modelo de regressão logística penalizada para o problema de diagnóstico de câncer a partir de dados de *microarray*; Xu *et al.*, (2016) utilizam um modelo RLM para análise de imagens; Ke *et al.*, (2006) propõem um modelo de predição de risco utilizando uma regressão logística multinomial de coeficientes semi-variáveis, entre outros.

Neste artigo, ajustamos um modelo de regressão logística multinomial para prever o nível de sombreamento em sistemas de criação de bovinos de corte. Apresentamos a função de verossimilhança de forma explícita, obtida com a introdução de variáveis latentes. Discutimos sobre o procedimento de estimação utilizando o pacote VGAM do software *R* (THOMAS, 2017). Ajustamos modelos com somente uma variável, com duas variáveis, com três variáveis e o modelo saturado com as quatro variáveis, totalizando o ajuste de quinze modelos.

Para selecionar o melhor modelo, consideramos os critérios de seleção de modelos AIC (AKAIKE, 1974) e BIC (SCHWARZ, 1978). O modelo selecionado é composto pelas variáveis ambientais, temperatura, luminosidade, umidade do ar e radiação UV. Verificamos a adequabilidade do modelo ajustado através do teste da razão de verossimilhanças. Apresentamos as medidas de ajuste Nagelkerk pseudo  $R^2$  (NAGELKERKE, 1991), McFadden (MCFADDEN, 1974) e Cox e Snell (COX AND SNELL, 1989) para justificar a adequabilidade do modelo ajustado. Além disso, apresentamos a tabela de classificação para o modelo ajustado; que apresenta uma taxa de 95,09% de acerto.

O restante do artigo está organizado da seguinte maneira. Na seção 2 apresentamos a plataforma eletrônica utilizada para obtenção das medidas ambientais e as estatísticas descritivas das medidas observadas. O ajuste dos quinze modelos de regressão logística multinomial é apresentado na Seção 3. Na Seção 4 descrevemos os resultados obtidos para os dados captados pela plataforma eletrônica. A Seção 5 apresenta as conclusões com base nos resultados obtidos.

## 2 Material e métodos

A plataforma desenvolvida pelo LSCAD, que consiste de sensores de captação de dados ambientais, é ilustrada na Figura 1.



Figura 1 - Plataforma de captação de dados ambientais.

Esta plataforma de aquisição de dados ambientais contém três sensores. O sensor DHT22, ilustrado na Figura 2(a), 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.

O resistor dependente de luz ou LDR, ilustrado na Figura 2(b), é responsável pela captação dos valores de luz ambiente. Seu alcance varia entre 0 e 1023, onde 0 representa menor resistência, ou seja, iluminação máxima e 1023 a ausência de luz.

O sensor responsável pela captação de raios UltraVioletas (UV) é o sensor UVM-30A, ilustrado na Figura 2(c). Seu alcance varia entre 0 e 11, onde 0 representa radiação baixa e 11 extrema. Para cada leitura feita pela plataforma também foi registrado o nível de sombreamento no momento da coleta.

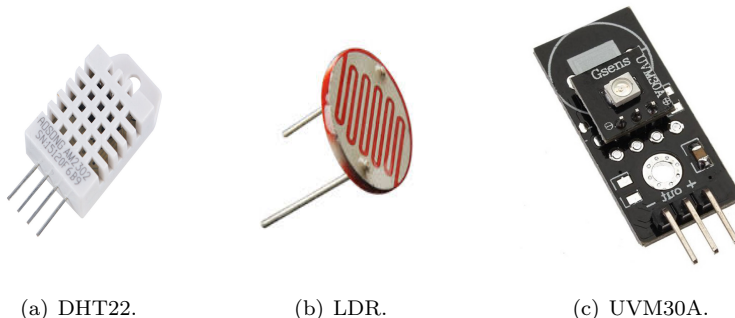


Figura 2 - Sensores responsáveis pela captação da umidade relativa do ar e temperatura ambiente (a), dos valores de luz ambiente (b) e dos raios UltraVioletas (c).

No restante do artigo, assumimos a seguinte notação para as variáveis ambientais:

- $X_1$  : Luminosidade,
- $X_2$  : Radiação Ultravioleta,
- $X_3$  : Temperatura,
- $X_4$  : Umidade relativa do ar.

Denotamos o nível de sombreamento por  $Y$  e assumimos a seguinte codificação:

$$Y = \begin{cases} 0, & \text{para sol,} \\ 1, & \text{para nublado,} \\ 2, & \text{para sombra.} \end{cases}$$

## 2.1 Modelo de regressão logística multinomial

Como a variável resposta  $Y$  (nível de sombreamento) assume um valor dentre três possíveis valores categoricos, consideramos um modelo de regressão logística multinomial com categoria de referência para os dados observados. Ou seja, assumimos que as probabilidades de  $Y$  assumir um dos valores do conjunto  $\{0, 1, 2\}$  são dadas por:

$$p_0(\boldsymbol{\beta}, \mathbf{x}) = P(Y = 0|\boldsymbol{\beta}, \mathbf{x}) = \frac{1}{1 + \exp\{\mathbf{x}'\boldsymbol{\beta}_1\} + \exp\{\mathbf{x}'\boldsymbol{\beta}_2\}} \quad (1)$$

$$p_1(\boldsymbol{\beta}, \mathbf{x}) = P(Y = 1|\boldsymbol{\beta}, \mathbf{x}) = \frac{\exp\{\mathbf{x}'\boldsymbol{\beta}_1\}}{1 + \exp\{\mathbf{x}'\boldsymbol{\beta}_1\} + \exp\{\mathbf{x}'\boldsymbol{\beta}_2\}} \quad (2)$$

$$p_2(\boldsymbol{\beta}, \mathbf{x}) = P(Y = 2|\boldsymbol{\beta}, \mathbf{x}) = \frac{\exp\{\mathbf{x}'\boldsymbol{\beta}_2\}}{1 + \exp\{\mathbf{x}'\boldsymbol{\beta}_1\} + \exp\{\mathbf{x}'\boldsymbol{\beta}_2\}} \quad (3)$$

com  $p_0(\boldsymbol{\beta}, \mathbf{x})$  sendo a categoria de referência, em que,  $\mathbf{x} = (1, x_1, x_2, x_3, x_4)$  e  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$  para  $\boldsymbol{\beta}_1 = (\beta_{01}, \beta_{11}, \beta_{21}, \beta_{31}, \beta_{41})$  e  $\boldsymbol{\beta}_2 = (\beta_{02}, \beta_{12}, \beta_{22}, \beta_{32}, \beta_{42})$  são os parâmetros do modelo.

Considerando o logaritmo da razão de chances em relação à categoria de referência, temos que

$$g_1(\boldsymbol{\beta}, \mathbf{x}) = \log\left(\frac{p_1(\boldsymbol{\beta}, \mathbf{x})}{p_0(\boldsymbol{\beta}, \mathbf{x})}\right) = \beta_{01} + \beta_{11}x_1 + \beta_{21}x_2 + \beta_{31}x_3 + \beta_{41}x_4, \quad (4)$$

e

$$g_2(\boldsymbol{\beta}, \mathbf{x}) = \log\left(\frac{p_2(\boldsymbol{\beta}, \mathbf{x})}{p_0(\boldsymbol{\beta}, \mathbf{x})}\right) = \beta_{02} + \beta_{12}x_1 + \beta_{22}x_2 + \beta_{32}x_3 + \beta_{42}x_4. \quad (5)$$

Considere associado a  $Y$  um vetor indicador  $Z = (Z_1, Z_2, Z_3)$ , tal que,  $Z$  assume uma das seguintes configurações:  $Z = (1, 0, 0)$  para  $Y = 1$ ,  $Z = (0, 1, 0)$  para  $Y = 2$  e  $Z = (0, 0, 1)$  para  $Y = 3$ . Dessa maneira, temos que  $Z \sim Multinomial(1, \mathbf{P})$ , onde  $\mathbf{P} = (p_0(\boldsymbol{\beta}, \mathbf{x}), p_1(\boldsymbol{\beta}, \mathbf{x}), p_2(\boldsymbol{\beta}, \mathbf{x}))$ .

Assim, dado uma amostra de tamanho  $n$ , a função de verossimilhança para  $\boldsymbol{\beta}$  é dada por

$$L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{x}, \mathbf{z}) = \prod_{i=1}^n \prod_{k=0}^2 p_{ik}(\boldsymbol{\beta}, \mathbf{x}_i)^{z_{ik}},$$

em que,  $\mathbf{y}$  é um vetor  $n \times 1$ ,  $\mathbf{x}$  é uma matriz  $n \times 5$  com  $\mathbf{x}_i$  sendo a  $i$ -ésima linha de  $\mathbf{x}$ ,  $\mathbf{z}$  é uma matriz indicadora  $n \times 3$  com  $\mathbf{z}_i = (z_{i1}, z_{i2}, z_{i3})$  sendo a  $i$ -ésima linha de  $\mathbf{z}$  e  $p_{ik}(\boldsymbol{\beta}, \mathbf{x}_i) = P(Y_i = k|\boldsymbol{\beta}, \mathbf{x})$ , para  $i = 1, \dots, n$  e  $k = 0, 1, 2$ .

A função log-verossimilhança é dada por

$$l(\boldsymbol{\beta}|\mathbf{y}, \mathbf{x}, \mathbf{z}) = \sum_{i=1}^n \sum_{k=0}^2 z_{ik} \log(p_{ik}(\boldsymbol{\beta}, \mathbf{x})) = \sum_{i=1}^n [z_{i1}\mathbf{x}'_i\boldsymbol{\beta}_1 + z_{i2}\mathbf{x}'_i\boldsymbol{\beta}_2 - \Psi(\boldsymbol{\beta}, \mathbf{x}_i)], \quad (6)$$

em que,  $\Psi(\boldsymbol{\beta}, \mathbf{x}_i) = \log(1 + \exp(\mathbf{x}'_i\boldsymbol{\beta}_1) + \exp(\mathbf{x}'_i\boldsymbol{\beta}_2))$ , para  $i = 1, \dots, n$ .

Derivando (6) em relação a  $\beta_{jk}$ , obtemos

$$\frac{dl(\boldsymbol{\beta}|\mathbf{y}, \mathbf{x}, \mathbf{z})}{d\beta_{jk}} = \sum_{i=1}^n z_{ik}x_{ij} - \frac{d\Psi(\boldsymbol{\beta}, \mathbf{x}_i)}{d\beta_{jk}} \quad (7)$$

para  $j = 0, 1, 2, 3, 4$  e  $k = 1, 2$ .

Os estimadores de máxima verossimilhança são obtidos igualando (7) a zero. Porém, devido ao termo  $\Psi(\boldsymbol{\beta}, \mathbf{x}_i)$ , não é possível obter os estimadores analiticamente. Devido a isto, as estimativas de máxima verossimilhança  $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$  serão obtidas numericamente. Para isto utilizamos o pacote *VGAM* e o comando *vglm* do software *R* (R CORE TEAM, 2018).

Dado as estimativas  $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$ , estimamos a probabilidade de  $Y = k$ , para  $k = 0, 1, 2$ , utilizando as probabilidades dadas em (1), (2) e (3) com  $\boldsymbol{\beta}_1 = \hat{\boldsymbol{\beta}}_1$  e  $\boldsymbol{\beta}_2 = \hat{\boldsymbol{\beta}}_2$ . Consideramos  $Y = k$  se  $k = \underset{m=0,1,2}{\operatorname{argmax}} P(Y = m|\hat{\boldsymbol{\beta}}, \mathbf{x})$ .

### 3 Aplicação

Nesta seção, apresentamos os resultados obtidos com o ajuste do modelo de regressão logística multinomial ao conjunto de dados obtido com a plataforma eletrônica. O conjunto de dados é formado por  $n = 698$  observações. Para cada leitura foi registrado o nível de sombreamento e os valores para as variáveis ambientais.

Como primeiro procedimento, dividimos as  $n = 698$  observações em dois conjuntos de dados. O conjunto de dados  $D_1$  é composto pelas primeiras 489 (70%) observações; e o conjunto de dados  $D_2$  é composto pelas 209 observações restantes. Esta divisão das  $n$  observações em dois conjuntos de dados foi feita com o objetivo de utilizar o conjunto de dados  $D_1$  para ajustar o modelo; e o conjunto de dados  $D_2$  para verificar a performance do modelo ajustado.

Como ilustração, apresentamos na Tabela 1, uma parte do conjunto de dados  $D_1$ , sendo duas observações para cada nível de sombreamento observado.

Tabela 1 - Parte do conjunto de dados  $D_1$

Observação	$X_1$	$X_2$	$X_3$	$X_4$	$Y$
11	3,0	5,0	31,5	58,7	0
30	3,0	6,0	32,1	59,3	0
59	4,0	4,0	34,2	52,6	1
112	6,0	2,0	30,2	72,3	1
80	20,0	0,0	34,0	49,6	2
273	9,0	0,0	34,0	54,8	2

Obtido o conjunto de dados  $D_1$ , nossa primeira análise consiste em calcular as estatísticas descritivas deste conjunto de dados. A Figura 3 mostra o gráfico em barras com as porcentagens observadas em cada uma das categorias. Das 489 observações, 45,60% foram leituras sob o estado 0 (“Sol”), 32,92% foram leituras sob o estado 1 (“Nublado”) e 21,47% foram leituras sob o estado 2 (“Sombra”).

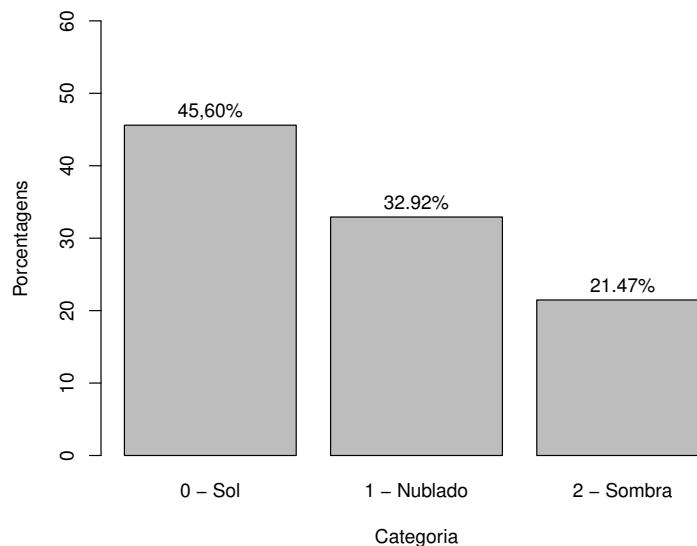


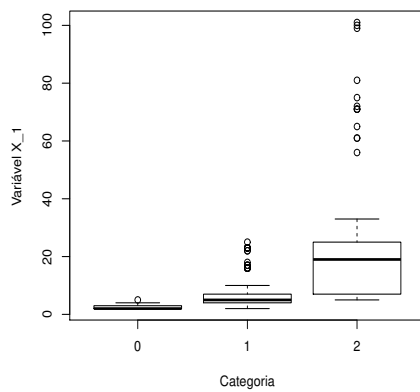
Figura 3 - Porcentagem observada de cada categoria.

A Tabela 2 mostra as estatísticas descritivas dos valores observados para cada variável ambiental. Para a variável  $X_1$ , luminosidade natural, o menor valor observado foi de 2 lux, o maior valor foi de 101 lux, com valor mediano de 4 lux e média e desvio-padrão de 8,268 e 13,152 lux, respectivamente. Já os valores observados para a variável  $X_2$ , radiação ultravioleta, variaram entre 0uv e 11uv, com valor mediano de 5uv, média de 4,524uv e desvio-padrão de 3,603uv. A variável  $X_3$ , temperatura, apresentou valor mínimo de 24,700°C, valor máximo de 49,9°C, valor mediano de 35,2°C e média e desvio-padrão de 36,460°C e 5,978°C, respectivamente. Para a variável  $X_4$ , umidade relativa do ar, os valores observados variaram de um mínimo de 18,20% e um máximo de 92,8% com valor mediano de 45,5%, média de 46,54% e desvio-padrão de 16,016%.

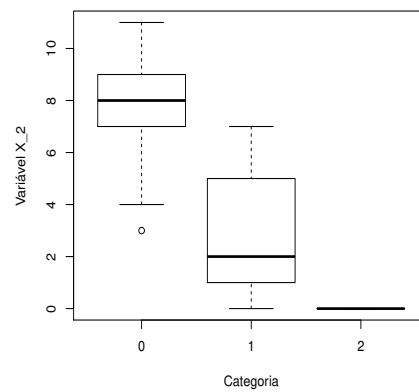
A Figura 4, mostra o *boxplot* dos valores observados para cada variável por valor categórico. Note que, a variável  $X_1$  apresenta os menores valores na categoria 0; e os maiores valores na categoria 2. A variável  $X_2$ , apresenta os maiores valores na categoria 0 (sol) e na categoria 2 todos os valores observados são iguais a 0 (zero). O valor mediano para a variável  $X_3$  reduz da categoria 0 para a categoria 2; enquanto que para a variável  $X_4$  o valor mediano aumenta.

Tabela 2 - Estatísticas descritivas

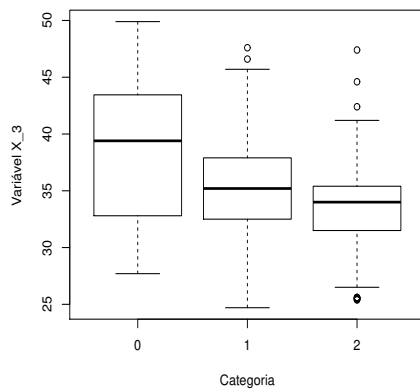
Variável	Min.	1º Q.	Mediana	Média	D.P.	3º Q.	Máx.
$X_1$	2,000	2,000	4,000	8,268	13,152	7,000	101,000
$X_2$	0,000	0,000	5,000	4,524	3,603	8,000	11,000
$X_3$	24,700	32,200	35,200	36,460	5,978	40,800	49,900
$X_4$	18,20	34,200	45,500	46,540	16,016	57,100	92,800



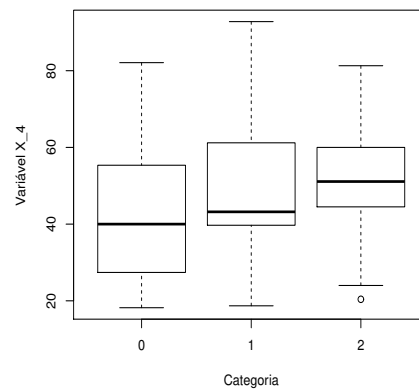
(a) Variável  $X_1$ .



(b) Variável  $X_2$ .



(c) Variável  $X_3$ .



(d) Variável  $X_4$ .

Figura 4 - Boxplot por variável e nível de sombreamento.



### 3.1 Ajuste do modelo

Utilizando o procedimento de estimação descrito na seção 2.1, ajustamos um modelo de regressão logística multinomial considerando as quatro variáveis ambientais e um modelo considerando somente o intercepto. Denotamos o modelo saturado por  $M_{1234}$  e o modelo com o intercepto por  $M_0$ . Então, realizamos o teste da razão de verossimilhanças para verificar se as variáveis  $X_j$ 's têm alguma influência sobre a variável resposta  $Y$ , para  $j = 1, 2, 3, 4$ .

Neste teste, a hipótese nula considera que nenhuma variável tem influência sobre a resposta  $Y$ , e a hipótese alternativa considera que pelo menos uma das variáveis tem influência sobre a variável respostas  $Y$ . A estatística de teste é dada por

$$D = -2\log \left( \frac{L(\hat{\beta}_0|\mathbf{y}, \mathbf{x}, \mathbf{z})}{L(\hat{\beta}|\mathbf{y}, \mathbf{x}, \mathbf{z})} \right),$$

onde  $\hat{\beta}_0 = (\hat{\beta}_{01}, \hat{\beta}_{02})$  são as estimativas dos interceptos do modelo  $M_0$ , com  $D$  tendo distribuição  $\chi^2$  com  $(k - 1)p$  graus de liberdade sob a hipótese nula, onde  $k$  é o número de categorias e  $p$  é o número de variáveis.

Realizamos o teste da razão de verossimilhanças assumindo um nível de significância  $\alpha = 0,05$ . A Tabela 3 apresenta o resultado do teste. Como o  $p$ -valor é menor do que o nível de significância  $\alpha$ , rejeitamos  $H_0$ . Ou seja, pelo menos uma das variáveis  $X_j$ ' é importante para explicar a variável resposta  $Y$ , para  $j = 1, 2, 3, 4$ .

Tabela 3 - Teste da razão de verossimilhanças

Modelo	-2 Log verossimilhança	Estatística $D$	Graus de Liberdade	$p$ -valor
$M_0$	1.030,9880	891,1977	8	< 0,00001
$M_{1234}$	139,7903			

Além do modelo saturado  $M_{1234}$ , ajustamos outros quatorze modelos, conforme descrito na Tabela 4. No total ajustamos quinze modelos. Este procedimento foi realizado com o objetivo de selecionar o modelo composto pelas variáveis exploratórias que melhor explicam a resposta categórica  $Y$ .

Para selecionar o melhor modelo dentre os ajustados, consideramos os critérios de seleção de modelos AIC e BIC, dados respectivamente, por

$$AIC_M = -2l(\hat{\beta}|\mathbf{y}, \mathbf{x}, \mathbf{z}) + 2H_M \text{ e } BIC_M = -2l(\hat{\beta}|\mathbf{y}, \mathbf{x}, \mathbf{z}) + H_M \log(n),$$

em que,  $l(\hat{\beta}|\mathbf{y}, \mathbf{x}, \mathbf{z})$  é a função log-verossimilhança com  $\hat{\beta}$  sendo as estimativas de máxima verossimilhança e  $H_M$  é a quantidade de parâmetros "livres" no modelo  $M$ . O modelo com o menor valor AIC ou BIC é o melhor modelo dentre os considerados.

A Tabela 5 apresenta os valores dos critérios AIC e BIC para cada modelo, em ordem crescente. Os valores destacados em negrito são os menores valores AIC e BIC. Assim, o modelo saturado  $M_{1234}$  é o melhor modelo dentre os considerados.

Tabela 4 - Modelos considerados.

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

Tabela 5 - Valores AIC e BIC para cada modelo.

Modelo	AIC	BIC	Modelo	AIC	BIC	Modelo	AIC	BIC
$M_{1234}$	<b>159,7900</b>	<b>201,7140</b>	$M_{23}$	259,1966	284,3508	$M_{13}$	459,1948	484,3484
$M_{234}$	174,9633	208,5022	$M_{12}$	284,6124	309,7666	$M_1$	467,1385	483,9079
$M_{124}$	185,1609	218,6998	$M_2$	302,8850	319,6245	$M_{34}$	933,4107	958,5648
$M_{24}$	208,1110	233,2652	$M_{134}$	428,6037	462,1426	$M_3$	958,7174	975,4869
$M_{123}$	211,4880	245,0269	$M_{14}$	451,1231	476,1773	$M_4$	1.000,7630	1.017,5330

A Tabela 6 mostra as estimativas e os intervalos de 95% de confiança para cada um dos parâmetros do modelo  $M_{1234}$ . Note que, o intervalo de confiança para  $\beta_{22}$  contém o valor 0. Ou seja, a variável  $X_2$  não é significativa para estimar  $Y = 2$ . Porém, esta variável é significativa para estimar  $Y = 1$ ; pois o intervalo para  $\beta_{21}$  não contém o valor zero. Devido a isto mantemos a variável  $X_2$  no modelo. Todos os outros intervalos não contém o valor 0, significando que as variáveis e os interceptos são significantes para predição de  $Y$ .

Tabela 6 - Estimativas para os parâmetros

Parâmetro	Estimativa	I.C. 95%	Parâmetro	Estimativa	I.C. 95%
$\beta_{01}$	62,1758	(33,4312; 90,9204)	$\beta_{02}$	59,9678	(16,7671; 103,1686)
$\beta_{11}$	1,5829	(0,7134; 2,4524)	$\beta_{12}$	1,6460	(0,7726; 2,5194)
$\beta_{21}$	-2,0283	(-2,7980; -1,2586)	$\beta_{22}$	-28,9544	(-880,7603; 842,8515)
$\beta_{31}$	-1,0541	(-1,5964; -0,5118)	$\beta_{32}$	-0,6926	(-1,3819; -0,0032)
$\beta_{41}$	-0,4052	(-0,5722; -0,2382)	$\beta_{42}$	-0,5344	(-0,7723; -0,2965)

Assim, as expressões em (4) e (5) são dadas por

$$g_1(\hat{\beta}, \mathbf{x}) = 62,1758 + 1,5829x_1 - 2,0283x_2 - 1,0541x_3 - 0,4052x_4;$$

$$g_2(\hat{\beta}, \mathbf{x}) = 59,9678 + 1,6460x_1 - 28,9544x_2 - 0,6926x_3 - 0,5344x_4.$$

A Tabela 7 mostra os valores do pseudo  $R^2$  para o modelo ajustado. Este valores representam a proporção da variabilidade dos dados que está sendo explicada pelo modelo ajustado. A estatística pseudo  $R^2$  de Nagelkerke indica que 95,43% da variabilidade está sendo explicada pelo modelo ajustado. As estatísticas pseudo  $R^2$  de McFadden e de Cox e Snell indicam 86,44% e 83,83%, respectivamente. Este valores foram obtidos utilizando o pacote *DescTools* e o comando *PseudoR2* do software *R*.

Tabela 7 - Valores de pseudo  $R^2$  para o modelo ajustado

Estatística	Nagelkerke	McFadden	Cox e Snell
Pseudo $R^2$	0,9543	0,8644	0,8383

Além das estatísticas pseudo  $R^2$ , também é importante verificar a acurácia do modelo ajustado. Isto é feito através da matriz de confusão. A Tabela 8 mostra a matriz de confusão para o modelo ajustado. Nesta Tabela, a quinta coluna mostra o percentual de classificação correta dentro de cada categoria. A quinta linha mostra o percentual predito em cada categoria. Como podemos notar, o modelo ajustado apresenta alta performance de predição correta. No geral, 95,09% de acerto.

Tabela 8 - Matriz de confusão do modelo ajustado

Observado	Predito			Percentual de classificação correta
	0	1	2	
0	212	11	0	95,07%
1	09	150	2	93,17%
2	0	2	103	98,10%
Percentual por categoria	45,20%	33,33%	21,47%	95,09%

### 3.2 Validação do modelo

Nesta seção descrevemos os resultados obtidos com a aplicação do modelo ajustado ao conjunto de dados  $D_2$ . Este conjunto de dados é composto por 209 observações, sendo 31,58% (66) de leituras sob o estado 0 (“Sol”), 67,46% (141) de leituras sob o estado 1 (“Nublado”) e 0,96% (2) de leituras sob o estado 2 (“Sombra”).

A Tabela 9 mostra a matriz de confusão para o conjunto de dados  $D_2$ . Dos valores observados como 0, 74,24% (49/66) foram corretamente preditos pelo modelo; enquanto que para os valores 1 observados, 99,29% (140/141) foram corretamente preditos; e os dois valores 2 observados foram corretamente preditos pelo modelo. No geral, o modelo ajustado apresenta 91,39% de acerto.

Tabela 9 - Matriz de confusão para o conjunto de dados  $D_2$

Observado	Predito			Percentual de classificação correta
	0	1	2	
0	49	17	0	74,24%
1	01	140	0	99,29%
2	0	0	2	100%
Percentual por categoria	23,92%	75,12%	0,96%	91,39%

Além do teste no conjunto de dados  $D_2$ , verificamos a capacidade de predição correta do modelo ajustado em outros  $L = 500$  conjuntos de dados, obtidos do conjunto de dados original com  $n = 698$  observações. Para isto, implementamos o seguinte procedimento de reamostragem. Para  $l = 1, \dots, L$ :

- (i) Selecione aleatoriamente  $n_t = 100$  linhas da matriz de dados original  $D$  e obtenha o conjunto de dados de teste  $D_l$ ;
- (ii) Aplique o modelo ajustado para predição dos valores  $Y$  de  $D_l$ ;
- (iii) Considere uma variável indicadora  $\mathbb{I}_i$  que assume o valor 1 se a  $i$ -ésima observação de  $D_l$  é corretamente predita; e  $\mathbb{I}_i = 0$  caso contrário, para  $i = 1, \dots, n_t$ ;
- (iv) Calcule o percentual de acerto  $\hat{P}_l = \frac{1}{n_t} \sum_{i=1}^{n_t} \mathbb{I}_i$ .

A Figura 5 mostra o *boxplot* para as proporções de acerto do modelo ajustado em 500 conjuntos de dados de teste. Em média, o modelo ajustado apresenta um percentual de acerto de 93,88%. Estes resultados mostram que o modelo ajustado apresenta uma satisfatória performance de predição correta.

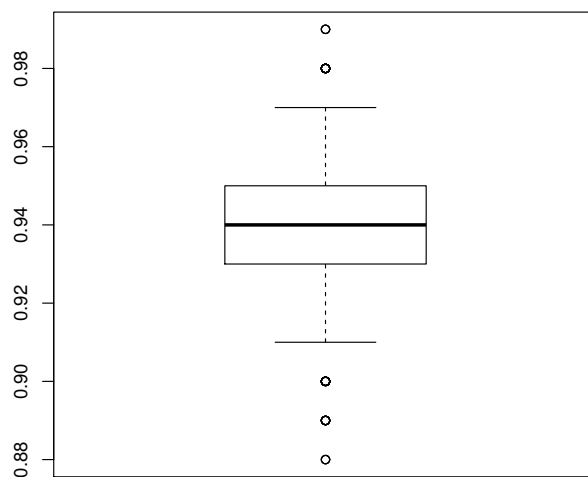


Figura 5 - Proporção de acerto em 500 conjuntos de dados de teste.

#### 4 Considerações finais

Neste artigo, apresentamos uma aplicação do modelo de regressão logística multinomial para a predição do nível de sombreamento em sistemas de criação de bovinos de corte. Os dados utilizados para o ajuste do modelo foram obtidos utilizando uma plataforma eletrônica composta por quatro sensores. As estimativas para os parâmetros do modelo foram obtidas numericamente via maximização da função de verossimilhança.

Para selecionar o modelo com as variáveis exploratórias que melhor explicam a variável resposta  $Y$  ajustamos quinze modelos. Selecionamos o melhor modelo utilizando os critérios de seleção de modelos AIC e BIC. O modelo selecionado foi o modelo composto pelas quatro variáveis: luminosidade, radiação UV, temperatura e umidade relativa do ar.

Verificamos a qualidade de ajuste do modelo através dos valores pseudo  $R^2$  e da matriz de confusão. Os valores pseudo  $R^2$  obtidos foram satisfatórios e a matriz de confusão mostrou uma alta capacidade de predição do modelo ajustado.

Do ponto de vista prático/econômico, a abordagem descrita neste artigo, composta pela plataforma eletrônica e o modelo de regressão logística ajustado, mostram-se promissores na identificação de animais que estão à procura de sombra, sendo este um indicativo de estresse térmico. Esta identificação é importante para que o responsável técnico (veterinário e/ou zootecnista) decida por uma intervenção com o objetivo de diminuir os efeitos do estresse térmico e propiciar uma melhor qualidade de vida ao animal. Reduzindo os efeitos do estresse térmico, segundo

Molento (2005), a carne produzida será de qualidade muito superior a de um animal que passou por algum tipo de manejo inadequado que provoque esse estresse.

Além disso, o método descrito neste artigo é uma contribuição importante na busca por novas estratégias de manejo para minimizar o efeito do calor. Com isso, diminuir o índice de doenças e aumentar o ganho de peso dos animais; o que consequentemente acarreta em uma redução dos custos e aumento de lucro do produtor.

Todas as análises estatísticas descritas foram implementadas no software R. Os códigos podem ser obtidos via email aos autores.

## Agradecimentos

Os autores agradecem o CNPq e a CAPES pelo apoio financeiro e aos revisores e editores pelas sugestões.

SANTOS, R. M. O.; SARAIVA, E. F.; ARCE, T. F.; SANTOS, R. R. Prediction of shading level using a multinomial logistic regression model in cutting cattle systems. *Rev. Bras. Biom.*, Lavras, v.37, n.3, p.378-393, 2019.

- **ABSTRACT:** *In the last few years, the consequences of thermal stress on beef cattle have been a subject of interest to researchers, investors and breeders. This is due to the fact that caloric stress has an impact on the decrease of the animals' weight gain and increases in slaughter time and cost of production, leading to the economic loss to the creators/investors. In this paper, we present an electronic platform developed to capture environmental measures in livestock systems. Using the observed data obtained by the electronic platform we fit fifteen Multinomial logistic regression models to predict the level of shading (Sun, Cloudy and Shadow) that a bovine is in. In order to select the best model we consider the model selection criteria AIC and BIC. As a result, we obtained a model with high predictive accuracy.*
- **KEYWORDS:** *Thermal stress; multinomial logistic regression; model selection; prediction.*

## Referências

- AGRESTI, A. *Categorical data analysis*. New York: John Wiley, 2012.
- 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.
- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v.19, p.716-723, 1974.
- COX, D. R.; SNELL, E. J. *Analysis of binary data*. 2.ed. London: Chapman & Hall, 1989.
- 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 p. Tese (Doutorado em Zootecnia), Universidade de São Paulo (USP), Pirassununga, 2008.
- KE, Y.; FU, B.; ZHANG, W. Semivarying coefficient multinomial logistic regression for disease progression risk prediction *Statistic in Medicine*, v.35, p.4764-4778, 2016.
- MARIN, C. *Sistema para aquisição e processamento de temperatura cutânea em bovinos e variáveis ambientais*, Dissertação (Mestrado), Universidade Federal de Mato Grosso do Sul, 2016.
- MITLÖHNER, F. M.; GALYEAN, M. L.; MACGLONE, J. J. Shade effects on performance, carcass traits, physiology, and behaviour of heat-stressed feedlot heifers. *Journal Animal Science*, v.80, p.2043-2050, 2002.
- MOLENTO, C. Bem-estar e produção animal: aspectos econômicos. *Archives of Veterinary Science*, v.10, 2005.
- NAGELKERKE, N. J. D. A Note on a general definition of the coefficient of determination. *Biometrika*, v.78, p.691-692, 1991.
- SILVA, R. G. *Introdução à bioclimatologia animal*. São Paulo: Nobel, 2000.
- R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL: <https://www.R-project.org/>.
- SCHWARZ, G. E. Estimating the dimension of a model. *Annals of Statistics*, v.6, p.461-464, 1978.
- THOMAS, W. Y. *VGAM: Vector generalized linear and additive models*. R package version 1.0-3, 2017. URL <https://CRAN.R-project.org/package=VGAM>
- VON BORRIES, G.; BASSINELLO, P. Z.; RIOS, E. S.; KOAKUZU, S. N.; CARVALHO, R. N. Prediction models of rice cooking quality. *Cereal Chemistry*, v.95, p.158-166, 2018.
- XU, Z.; HONG, Z.; ZHANG, Y.; WU, J.; TSOI, A. C.; TAO, D. Multinomial latent logistic regression for image understanding. *IEEE Transactions on Image Processing*, v.25, p.973-987, 2016.

ZHU, J.; HASTIE, T. Classification of gene microarrays by penalized logistic regression. *Biostatistics*, v.5, p.427-443, 2004.

Recebido em 29.02.2018.

Aprovado após revisão em 26.04.2019.