



LUIZ FERNANDO SILVA RESENDE

**MÉTODO DE SEGMENTAÇÃO POR BOXPLOT
CIRCULAR PARA DADOS MULTIMODAIS E SUA
APLICAÇÃO NA ANÁLISE DE CAPTURA DE ENXAMES
DE ABELHAS COM EVENTOS RECORRENTES**

LAVRAS – MG

2025

LUIZ FERNANDO SILVA RESENDE

**MÉTODO DE SEGMENTAÇÃO POR BOXPLOT
CIRCULAR PARA DADOS MULTIMODAIS E SUA
APLICAÇÃO NA ANÁLISE DE CAPTURA DE ENXAMES
DE ABELHAS COM EVENTOS RECORRENTES**

Tese apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para obtenção do título de Doutor.

Profa. Dra. Carla Regina Guimarães Brighenti
Orientadora

**LAVRAS – MG
2025**

**Ficha Catalográfica elaborada pelo Sistema de Geração
de Ficha Catalográfica da Biblioteca Universitária da UFLA, com
dados informados pelo(a) próprio(a) autor(a).**

Resende, Luiz Fernando Silva.

Método de segmentação por boxplot circular para dados multimodais e sua aplicação na análise de captura de enxames de abelhas com eventos recorrentes / Luiz Fernando Silva Resende. - 2025.

159 p. : il.

Orientadora: Carla Regina Guimarães Brighenti

Tese (Doutorado) - Universidade Federal de Lavras, 2025.

Bibliografia.

1. Modelos de Andersen-Gill. 2. Boxplot circular. 3. Abelhas africanizadas. 4. Estatística circular. 5. Segmentação de dados. I. Brighenti, Carla Regina Guimarães. II. Universidade Federal de Lavras. III. Título.

LUIZ FERNANDO SILVA RESENDE

**MÉTODO DE SEGMENTAÇÃO POR BOXPLOT
CIRCULAR PARA DADOS MULTIMODAIS E SUA
APLICAÇÃO NA ANÁLISE DE CAPTURA DE ENXAMES
DE ABELHAS COM EVENTOS RECORRENTES**

**CIRCULAR BOXPLOT SEGMENTATION METHOD FOR MULTIMODAL DATA
AND ITS APPLICATION IN THE ANALYSIS OF BEE SWARM CAPTURE WITH
RECURRENT EVENTS**

Tese apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para obtenção do título de Doutor.

APROVADA em 29 de abril de 2025.

Profa. Dra. Luciane Teixeira Passos Giarola	UFSJ
Prof. Dr. José Waldemar da Silva	UFU
Prof. Dr. Deodoro Magno Brighenti dos Santos	UFSJ
Prof. Dr. Clodoaldo Teodosio Santana da Silva	UFVJM

Profa. Dra. Carla Regina Guimarães Brighenti
Orientadora

**LAVRAS – MG
2025**

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Agradeço à CAPES pelo suporte financeiro essencial para o desenvolvimento desta pesquisa. Expresso minha profunda gratidão à minha orientadora, Profa. Dra. Carla Regina Guimarães Brighenti, pelo acompanhamento dedicado e pelas valiosas contribuições ao longo do doutorado. Agradeço também aos membros da banca, Profa. Dra. Luciane Teixeira Passos Giarola, Prof. Dr. José Waldemar da Silva, Prof. Dr. Deodoro Magno Brighenti dos Santos e Prof. Dr. Clodoaldo Teodosio Santana da Silva, pelas sugestões e avaliações que enriqueceram este trabalho. À Universidade Federal de São João del-Rei (UFSJ), meu reconhecimento pelo espaço cedido no campus CTAN e pelos materiais disponibilizados para a defesa e execução do experimento, bem como ao Grupo de Estudos em Apicultura (GEAPIS) da UFSJ pelo apoio técnico e logístico fundamental na realização do experimento de captura de enxames de abelhas africanizadas, que viabilizou a coleta de dados e o desenvolvimento do Artigo 3 e 4 desta tese. Agradeço à Universidade Federal de Lavras (UFLA) pela infraestrutura acadêmica e suporte institucional durante meu mestrado e doutorado, bem como aos professores do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, que contribuíram para minha formação. Por fim, meu carinho e gratidão aos meus pais, Luiz Humberto e Tânia, cujo amor inabalável, incentivo constante e sacrifícios diários moldaram quem sou hoje, inspirando-me a persistir em cada desafio acadêmico. E à minha namorada Nayara, cujo apoio incondicional e auxílio na execução do experimento foram essenciais para a conclusão desta jornada.

RESUMO

A estatística circular é aplicada em diversos ramos da ciência, principalmente quando os dados envolvem sazonalidade, horários e direções. A presença de multimodalidade em conjuntos de dados é um desafio na análise e modelagem estatística. Esse problema também é comum em dados circulares, em que a segmentação de cada conjunto modal é realizada de forma empírica, utilizando, por exemplo, dias da semana ou estações. Outra alternativa é o uso de algoritmos de maximização a partir da modelagem por misturas de distribuições von Mises, que exige mais conhecimento teórico. Essa tese propõe, no primeiro artigo, um método de segmentação a partir da construção de boxplot circular. O método foi apresentado e sua eficácia verificada através de simulação em diferentes cenários e posteriormente aplicado a dados sobre horários de acidentes de trânsito que apresentaram multimodalidade ao longo da semana. Verificou-se que o método proposto foi mais adequado para o estudo de multimodalidade em dados circulares ao equilibrar simplicidade e precisão. No terceiro artigo, realizou-se um experimento para caracterizar a captura de enxames de abelhas africanizadas, durante um ano, a partir da instalação de ninhos-armadilha, registrando-se dados lineares, tais como altura, altitude, precipitação, umidade e velocidade do vento, bem como as circulares: direção do vento, época de captura (mês), longitude, latitude e ângulo do alvado, que representa o posicionamento do ninho em relação ao Norte magnético. Além disso, avaliou-se a variável qualitativa cor de revestimento do ninho-armadilha. A época de captura apresentou padrões bimodais que foram estudados a partir do Método de Segmentação por Boxplot Circular desenvolvido no Artigo 1. Verificou-se a relação entre as variáveis registradas no experimento através da correlação linear-linear, linear-circular e circular-circular, de acordo com a natureza de cada variável. No quarto artigo, devido à presença de dados censurados, já que as datas de captura foram registradas em intervalos, entre o último dia com ausência do enxame e o dia da observação da captura, foi utilizada a análise de sobrevivência com o modelo de Andersen-Gill, considerando covariáveis lineares e circulares para modelar o tempo até a captura de enxames, com segmentação bimodal via boxplot circular. Além disso, foi considerada a segmentação dos dois períodos sazonais, pós-chuva, entre 10 de abril a 06 de junho, e pré-chuva, de 29 de agosto a 19 de outubro, na análise dos dados a fim de considerar a bimodalidade apresentada na caracterização da captura.

Palavras-chave: boxplot circular; distribuição von Mises; correlação circular; acidentes de trânsito; captura de abelhas.

ABSTRACT

Circular statistics is applied across various scientific fields, particularly when data involve seasonality, time, or directions. Multimodality in datasets poses a significant challenge for statistical analysis and modeling, a common issue in circular data where modal subsets are often segmented empirically, such as by days of the week or seasons. Alternatively, maximization algorithms based on mixtures of von Mises distributions are used, requiring advanced theoretical knowledge. This thesis, structured in four articles, proposes innovative approaches to address these challenges. The first article introduces a segmentation method based on the construction of a circular boxplot, with its effectiveness validated through simulations across diverse scenarios and applied to multimodal traffic accident timing data throughout the week, demonstrating a balance between simplicity and accuracy. The second article extends this methodology, refining the circular boxplot approach for enhanced precision in multimodal data analysis. The third article presents an experiment characterizing Africanized bee swarm capture over one year using trap nests, recording linear variables (height, altitude, precipitation, humidity, wind speed) and circular variables (wind direction, capture season, longitude, latitude, nest entrance angle relative to magnetic North), alongside the qualitative variable of trap nest coating color. Bimodal patterns in capture seasons were analyzed using the circular boxplot segmentation method developed in the first article, with relationships between variables explored through linear-linear, linear-circular, and circular-circular correlations based on their nature. The fourth article addresses censored data, as capture dates were recorded within intervals between the last day without a swarm and the observation day, employing survival analysis with the Andersen-Gill model, incorporating linear and circular covariates to model swarm capture time, with bimodal segmentation via circular boxplot. Additionally, two seasonal periods—post-rain (April 10 to June 6) and pre-rain (August 29 to October 19)—were segmented to account for the bimodal patterns observed in swarm capture characterization. The proposed methods and findings offer practical and accessible tools for analyzing multimodal circular data, with applications in traffic safety and beekeeping, supporting biodiversity and food security through enhanced swarm capture efficiency.

Keywords: circular boxplot; von Mises distribution; circular correlation; traffic accidents; bee swarm capture.

INDICADORES DE IMPACTO

Esta tese, centrada no desenvolvimento do Método de Segmentação por Boxplot Circular e sua aplicação na análise de captura de enxames de abelhas africanizadas, apresenta impactos potenciais nas dimensões social, tecnológica, econômica e ambiental, com relevância para comunidades rurais de Minas Gerais, especialmente nas regiões apícolas do Sul de Minas e Campo das Vertentes. A metodologia proposta, que equilibra simplicidade e precisão na análise de dados circulares multimodais, pode otimizar a eficiência da captura de enxames, fortalecendo a apicultura e contribuindo para a segurança alimentar por meio da polinização e produção de mel, além de apoiar a conservação de abelhas, essenciais para a biodiversidade. Na segurança de trânsito, a segmentação de horários de acidentes pode subsidiar políticas públicas, promovendo cidades mais seguras.

IMPACT INDICATORS

This thesis, focused on the development of the Circular Boxplot Segmentation Method and its application in the analysis of Africanized bee swarm capture, presents potential impacts in social, technological, economic, and environmental dimensions, with relevance to rural communities in Minas Gerais, particularly in the beekeeping regions of Sul de Minas and Campo das Vertentes. The proposed methodology, which balances simplicity and accuracy in analyzing multimodal circular data, can optimize swarm capture efficiency, strengthening beekeeping and contributing to food security through pollination and honey production, while supporting bee conservation, essential for biodiversity. In traffic safety, time-based segmentation of accidents can support public policies, promoting safer cities.

LISTA DE FIGURAS

Figura 2.1 – Ponto no plano.	20
Figura 2.2 – Representação visual da mediana circular com amostras de tamanhos (a) ímpar e (b) par.	21
Figura 2.3 – Boxplot para o caso (a) linear e (b) circular.	22
Figura 2.4 – Gráficos para dados circulares: (a) diagrama de dispersão circular, (b) histograma circular, (c) diagrama de rosa.	23
Figura 2.5 – Distribuições no círculo: (a) uniforme, (b) unimodal e (c) 4-modais simétricas.	26
Figura 2.6 – Representação de dados bimodais (a) lineares e (b) circulares.	26
Figura 2.7 – Representação gráfica de registros de tempos com censura e sem censura. Legenda: (- -) pontilhado vertical representa o início e fim do estudo; (—) retas horizontais corresponde ao período que o indivíduo esteve sob risco; (●) a ocorrência do evento; (○) censura à direita; (▽) censura à esquerda; (- · -) pontilhado horizontal representa a censura intervalar com (■) sendo os intervalos de início e fim.	33
Figura 2.8 – Representação gráfica de registros dos tempos: (a) Total Time, (b) Counting Process e (c) Gap Time.	35

LISTA DE SÍMBOLOS

θ	Observação angular em uma amostra de tamanho n de uma variável aleatória circular
n	Tamanho da amostra de uma variável aleatória
i	Índice utilizado para contar as observações em uma amostra de tamanho n
$\bar{\theta}$	Média circular
C	Cosseno do vetor médio na equação da média circular
S	Seno do vetor médio na equação da média circular
$\tilde{\theta}$	Mediana circular
A	Ponto da mediana circular $\tilde{\theta}$ no diâmetro AB
B	Ponto oposto a A no diâmetro AB
Q_H	Quartil no sentido horário à mediana circular $\tilde{\theta}$
Q_A	Quartil no sentido anti-horário à mediana circular $\tilde{\theta}$
L_H	Limite horário no boxplot circular
L_A	Limite anti-horário no boxplot circular
f	Função densidade de probabilidade (PDF)
U	Distribuição circular uniforme
vM	Distribuição de von Mises
μ	Parâmetro da média na distribuição de von Mises
κ	Parâmetro de concentração na distribuição de von Mises
$I_0(\kappa)$	Função de Bessel modificada de ordem zero na distribuição de von Mises
M	Número de modas em uma distribuição multimodal circular
P_{nk}	Probabilidade do ponto de dado x_n pertencer ao cluster com índice k no modelo EM
x_n	Uma observação em θ
K	Número de componentes no modelo de mistura von Mises
ω_k	Proporção de cada componente com índice k no modelo de mistura von Mises
μ_k	Parâmetro de média direcional para o componente com índice k no modelo de mistura von Mises

κ_k	Parâmetro de concentração para o componente com índice k no modelo de mistura von Mises
$\mathcal{L}(\mu, \kappa, \pi)$	Função de verossimilhança a ser maximizada no algoritmo EM
X	Variável linear genérica
Y	Segunda variável linear genérica
$\bar{x}$	Média amostral da variável linear X
$\bar{y}$	Média amostral da variável linear Y
r_{xy}	Coefficiente de correlação de Pearson entre as variáveis lineares X e Y
ϕ	Variável angular em um círculo unitário
$\bar{\phi}$	Média direcional da variável angular ϕ
V	Coefficiente de Cramer para associação entre variáveis qualitativas
χ^2	Estatística do teste Qui-Quadrado de Pearson
m	Número de linhas da tabela de contingência
T	Tempo de sobrevivência
$Y_i(t)$	Variável indicadora de risco do indivíduo i no instante t
T_0	Último momento sem evento na censura intervalar
T_F	Momento em que o evento é verificado na censura intervalar
t_{ij}	Tempo da j -ésima ocorrência do evento para o i -ésimo sujeito
j	Índice para o j -ésimo grupo ou conjunto de dados
g_{ij}	Intervalo entre ocorrências consecutivas do sujeito i
$\lambda(t)$	Função de risco ou taxa de falha
$\lambda_0(t)$	Risco basal no modelo de Cox
x_i	Covariáveis do sujeito i no modelo de Cox
β	Coefficiente das covariáveis no modelo de Cox

SUMÁRIO

	PRIMEIRA PARTE	13
1	INTRODUÇÃO	14
2	REFERENCIAL TEÓRICO	20
2.1	Dados circulares	20
2.1.1	Estatística descritiva para dados circulares	21
2.1.2	Gráficos para Dados Circulares	22
2.1.3	Modelos probabilísticos para dados circulares	23
2.1.3.1	Distribuição circular uniforme	24
2.1.3.2	Distribuição von Mises	24
2.2	Multimodalidade na análise de dados	25
2.2.1	Métodos de segmentação em dados circulares multimodais	27
2.2.1.1	Método Empírico	27
2.2.1.2	Método Probabilístico por algoritmo EM	28
2.2.2	CrITÉrios de Informação AIC e BIC	30
2.3	Análise de correlação entre variáveis	30
2.4	Dados de Sobrevivência	32
2.4.1	Mecanismo de Censura	33
2.4.2	Eventos Recorrentes	34
2.4.3	Modelos na Análise de Sobrevivência	35
2.4.4	Modelo Semiparamétrico de Cox	36
2.4.5	Extensão do Modelo de Cox para dados recorrentes: Modelo de Andersen- Gill	36
	SEGUNDA PARTE	38
	Artigo 1 - Método de segmentação por boxplot circular	39
	Artigo 2 - Segmentação dos dias da semana em dados de acidentes de trân- sito nas rodovias mineiras	72
	Artigo 3 - Influência de fatores ambientais e sazonais em São João del Rei na captura de enxames de abelhas africanizadas	93
	Artigo 4 - Análise de captura de enxames de abelhas africanizadas com eventos recorrentes	123
	TERCEIRA PARTE	157

3	CONSIDERAÇÕES FINAIS	158
	REFERÊNCIAS	159

PRIMEIRA PARTE

1 INTRODUÇÃO

Dados circulares surgem naturalmente em muitos campos científicos nos quais as observações são registradas como ângulos ou direções em relação a um sistema com uma orientação fixa. Uma possibilidade de dados circulares surge da conversão de dados temporais periódicos em medidas angulares quando estes são redimensionados a uma circunferência de tamanho 2π , correspondente a um círculo unitário de raio 1 (Wang, 2013).

A aplicação da estatística circular em dados temporais permite examinar padrões que podem se manifestar em diferentes escalas, como diários, semanais, mensais, ou anuais, dependendo do contexto dos dados (Brunsdon; Corcoran, 2006). A identificação e análise de tais padrões são fundamentais para entender a distribuição dos pontos na circunferência. Dessa forma, o cálculo de estatísticas de posição e dispersão circulares, a construção de gráficos como o diagrama de dispersão circular e o histograma circular, bem como o uso de modelos de distribuição circular, permitem verificar tendências e variações nos dados que não seriam aparentes com análises tradicionais (Pewsey; García-Portugués, 2021).

Outra importante ferramenta que permite estudar dados circulares temporais com maior precisão é o boxplot circular. Seu estudo é relativamente novo, com uma adaptação da técnica linear para dados circulares sendo construída inicialmente por Anderson (1993) e, mais recentemente, Buttarazzi et al. (2018) propuseram uma nova abordagem para o boxplot que é a metodologia utilizada no pacote `bpDir` (Buttarazzi, 2022) do software R (Team, 2025).

Destaca-se também que, assim como em dados lineares, os dados circulares também podem apresentar mais de uma moda. Segundo Chakraborty (2014), dados observados, em situações da vida real podem apresentar mais que uma moda, isto é, são multimodais ou possuem vários picos.

Detectar uni ou multimodalidade é muito útil em várias aplicações de análise de dados (Chasani; Likas, 2024). A unimodalidade está diretamente relacionada ao comportamento de agrupamento dos pontos, ou seja, se os dados estão reunidos ou não. O exemplo mais típico de unimodalidade em dados circulares são da distribuição von Mises (1918), que é uma generalização da distribuição normal para dados circulares, amplamente utilizada, quando os dados circulares são agrupados em torno da média circular μ e com concentração κ (Mardia; Jupp, 2000).

A distribuição de von Mises é amplamente utilizada na modelagem de variáveis aleatórias circulares por sua flexibilidade e adequação a diversos contextos estatísticos. Um caso particular dessa distribuição ocorre quando o parâmetro de concentração é $\kappa = 0$, resultando na distribuição uniforme circular (Donner; Goldenshluger, 2024). Nesse cenário, a função densidade de probabilidade assume um valor constante, dado por $\frac{1}{2\pi}$, indicando que todos os ângulos possuem a mesma probabilidade de ocorrência ao longo do círculo unitário.

Embora a distribuição von Mises seja adequada para modelar dados circulares unimodais, muitos conjuntos de dados do mundo real apresentam comportamento multimodal, nos quais as observações se agrupam em torno de múltiplas direções ou pontos temporais distintos. Nesses casos, assumir uma única distribuição von Mises pode não capturar a estrutura subjacente dos dados. Para abordar isso, pesquisadores frequentemente utilizam técnicas de agrupamento ou modelos de mistura, como misturas de distribuições von Mises, para identificar e caracterizar múltiplas modas (Marković; Petrović, 2012). Essas abordagens permitem a segmentação de dados circulares multimodais em grupos distintos, cada um potencialmente seguindo sua própria distribuição von Mises com parâmetros únicos. Essa mudança de modelagem unimodal para multimodal motiva a exploração de algoritmos de agrupamento avançados adaptados para dados circulares, que podem detectar e separar eficientemente essas estruturas modais.

Chang-Chien et al. (2012) introduziram um algoritmo de agrupamento, baseado no deslocamento médio (MSBC), adaptado especificamente para dados circulares para determinar automaticamente o número de clusters e identificar centros de agrupamento em dados multimodais. Kesemen et al. (2016) propuseram o algoritmo de agrupamento de médias c-fuzzy, para dados direcionais, que permitiu resultados superiores em termos de consistência, precisão e tempo computacional, em comparação com outros algoritmos de agrupamento.

E, para os casos em que se observa multimodalidade, o Modelo de Mistura von Mises, que é uma extensão da distribuição von Mises, é ideal para descrever esses dados que podem ser agrupados. O modelo atribui a cada agrupamento uma distribuição von Mises com seu próprio conjunto de parâmetros, média circular μ e concentração κ . A proporção de cada um desses agrupamentos é determinada pelo parâmetro ω . A capacidade de modelar múltiplos agrupamentos com parâmetros distintos permite uma análise mais detalhada e precisa dos dados direcionais (Banerjee *et al.*, 2005). Para estudar esses modelos, Rodríguez et al. (2020)

propuseram uma abordagem Bayesiana para modelagem de misturas na análise de clusters de dados circulares.

Os modelos de mistura são uma ferramenta popular e amplamente utilizados para agrupamento em virtude de seus fundamentos probabilísticos e flexibilidade (Bouveyron; Brunet-Saumard, 2014). Nesse método, os dados são atribuídos aos agrupamentos, com base na distribuição do componente com maior probabilidade de tê-lo gerado. O procedimento de agrupamento envolve ajustar um modelo de mistura, frequentemente usando o algoritmo de Maximização de Expectativa (EM), e atribuir cada ponto de dado ao componente com a maior probabilidade posterior (Chasani; Likas, 2024).

A complexidade das técnicas de agrupamento para dados circulares dificulta o uso por pesquisadores que não possuem conhecimento aprofundado nesse tema, que recorrem ao determinismo empírico atribuído aos dados, que é mais simples e intuitivo. Esse Método Empírico consiste em uma abordagem na qual o início e o fim do ciclo são determinados de maneira arbitrária, não levando em consideração as características dos dados no ciclo. Como exemplo, pode-se citar o uso das datas do ano civil no caso de ciclos anuais, ou seja, ciclos iniciando em janeiro e terminando em dezembro ou a atribuição de pontos de segmentação para os agrupamentos utilizando convenções cotidianas, como definir 23h59min como o fim de um agrupamento e 00h00 como o início de outro. Banerjee et al. (2005) apresentam mistura de distribuições circulares de von Mises na identificação de agrupamentos que seriam difíceis de detectar com métodos empíricos.

Existem também ferramentas gráficas apropriadas para visualização de distribuições multimodais tais como o histograma circular e linhas de densidade. Já o boxplot circular, bem como o boxplot tradicional, é planejado para distribuições unimodais (Buttarazzi; Pandolfo; Porzio, 2018). No entanto, o boxplot para dados lineares pode ser adaptado para a análise de dados multimodais por meio de técnicas avançadas, como demonstrado por Qarmalah et al. (2016). Esses autores introduziram o conceito de *k-boxplots*, para permitir a visualização clara e eficaz de múltiplas modas em dados de mistura.

O Artigo 1 deste trabalho, propõe um método de segmentação a partir do deslocamento dos limites empíricos, que geralmente iniciam em 0° . O deslocamento será realizado a partir da construção do boxplot circular que não demanda um conhecimento profundo, somente o cálculo de quartis em dados circulares. Compara-se os resultados com o Método Empírico e com o mais utilizado, o Método Probabilístico por algoritmo EM, a partir de misturas de distribuições von

Mises. As possibilidades do método proposto incluem a análise de dados sazonais, horários e direções, com aplicações em campos como ecologia, meteorologia e segurança viária, onde a multimodalidade é comum. O Método de Segmentação por Boxplot Circular (MSBC) equilibra simplicidade e precisão, facilitando a identificação de padrões cíclicos sem a necessidade de algoritmos complexos, tornando-o acessível para pesquisadores de diversas áreas.

Uma aplicação relevante do MSBC é na análise de dados de acidentes de trânsito, onde os horários de ocorrência exibem padrões multimodais ao longo da semana, como picos matinais e noturnos. Ao segmentar esses dados, o método permite identificar variações semanais, subsidiando políticas de segurança pública, como alocação de recursos policiais em horários críticos. Essa abordagem, referenciada no Artigo 2, demonstra a eficácia do MSBC em equilibrar simplicidade e precisão comparado a métodos empíricos ou probabilísticos.

O MSBC, ao lidar com multimodalidade em dados sazonais e direcionais, estende-se naturalmente a aplicações ecológicas, como o estudo do comportamento de abelhas, onde padrões cíclicos influenciam processos como nidificação e migração. A nidificação de abelhas ocorre em locais variados, como árvores, solo ou rachaduras em estruturas. Estudos com meliponíneos, abelhas sem ferrão, indicam que 85% preferem cavidades em árvores vivas, com menores percentuais em árvores mortas, troncos secos ou cupinzeiros (Macedo *et al.*, 2020).

Além de cavidades adequadas e floradas favoráveis, a orientação magnética dos ninhos é crucial, influenciando a direção dos orifícios de entrada (Macedo *et al.*, 2020; Lacerda *et al.*, 2017; Vaz *et al.*, 2021). Análises revelam nanopartículas de magnetita no abdômen de abelhas *Apis mellifera* e sem ferrão, sugerindo mecanismos de orientação magnética (Vaz *et al.*, 2021; Gould; Kirschvink; Deffeyes, 1978; Lucano, 2006; Esquivel *et al.*, 2007). A navegação das abelhas combina pistas visuais memorizadas com a posição do sol, atuando como um "relógio solar" (Capaldi; Dyer, 1999; Beier; Lindauer, 1970).

Condições climáticas, como temperatura e umidade, afetam diretamente as atividades das abelhas africanizadas, influenciando produção de mel, coleta de recursos e resistência a parasitas (Silva *et al.*, 2023; Sousa; Rizzardo, 2014; Pinto; Silva; Anjos, 2024). Em períodos de enxameação, motivados por escassez ou excesso de recursos, apicultores instalam ninhos-armadilha para capturar enxames de forma sustentável e de baixo custo (Couvillon; Schürch; Ratnieks, 2014; Leven *et al.*, 2006; Vinícius-Silva; Ferraz; Werneck, 2021). Fatores como altura, localização, posicionamento solar e climáticos afetam o sucesso da captura (Qu *et al.*, 2022; Steffan-Dewenter; Kuhn, 2003).

Muitas variáveis dependem da época do ano, podendo gerar multimodalidade nos eventos de captura. Assim, o Artigo 3 relata um experimento realizado para captura de abelhas africanizadas, visando estudar variáveis qualitativas (cor do ninho) e quantitativas lineares (altura, precipitação, umidade, velocidade do vento) e circulares (ângulo do alvado, direção do vento, época), empregando correlações adaptadas e aplicando o MSBC à variável circular época da captura de enxame no ano, para identificar multimodalidade sazonal, fornecendo recomendações práticas para apicultores.

Após caracterizar os padrões multimodais na captura de enxames, surge a necessidade de modelar o tempo até esses eventos, considerando recorrências e censura nos dados, o que motiva a integração de análise de sobrevivência com variáveis circulares. A análise de sobrevivência estuda o tempo até um evento de interesse, lidando com dados censurados parciais ou incompletos (Turkson; Ayiah-Mensah; Nimoh, 2021), aplicada em medicina, engenharia, trânsito e ecologia para modelar tempos influenciados por múltiplos fatores (Liang *et al.*, 2022; RASHEED, 2023; Tesfay *et al.*, 2025).

Na apicultura, é adequada para padrões temporais de captura de enxames, com interações biológicas, ambientais e geográficas (Brodschneider; Gray, 2022). Capturas envolvem recorrências, com reposição de ninhos vazios caracterizando eventos recorrentes dependentes temporalmente. O modelo Andersen-Gill estuda tempos recorrentes, como em surtos de doenças (Farooq *et al.*, 2025). O Artigo 4 integra estatística circular à análise de sobrevivência via modelo Andersen-Gill, avaliando recaptura e covariáveis influenciando o tempo até captura de enxames africanizados.

Este trabalho está estruturado da seguinte maneira: na PRIMEIRA PARTE foi apresentada uma Introdução com uma breve exposição sobre multimodalidade e estudos em dados circulares. Na sequência, a seção de Notações e Definições exhibe os principais itens e símbolos necessários para a compreensão do método proposto e suas aplicações nos demais artigos.

Na SEGUNDA PARTE serão apresentados 4 artigos:

- a) artigo 1: Método de Segmentação do Boxplot Circular (MSBC);
- b) artigo 2: Aplicação do MSBC em dados de acidentes de trânsito;
- c) artigo 3: Caracterização da Captura de Enxames de Abelhas Africanizadas: Análise de Variáveis Lineares e Circulares;
- d) artigo 4: Análise de Sobrevivência na Captura de Enxames de Abelhas Africanizadas com Eventos Recorrentes.

Na TERCEIRA PARTE, as considerações finais sintetizam os resultados e discutem as implicações.

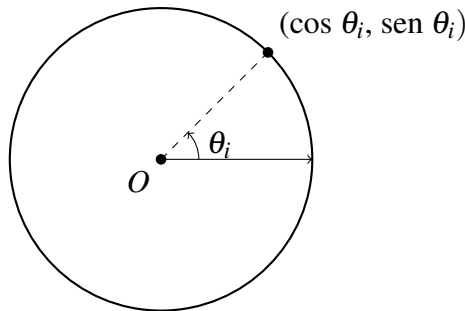
2 REFERENCIAL TEÓRICO

Inicialmente, serão expostas as principais definições necessárias para apresentar e esclarecer o método proposto.

2.1 Dados circulares

Na estatística circular, uma amostra de observações angulares θ de tamanho n , denotada por $\theta_1, \theta_2, \dots, \theta_n$, pode ser representada geometricamente por pontos, em uma circunferência centrada na origem do sistema de coordenadas cartesianas, com raio $r = 1$, em que cada observação θ_i com índice i corresponde ao par ordenado $(\cos \theta_i, \sin \theta_i)$, como ilustrado na Figura 2.1. Essa representação facilita o cálculo de diversas estatísticas, como medidas de posição e dispersão, que levam em consideração a natureza cíclica dos dados (Pewsey, 2004).

Figura 2.1 – Ponto no plano.



Fonte: Adaptado de MORALES (2006).

A posição direcional possui representação única num sistema coordenado de duas dimensões, isto é, qualquer ponto sobre o plano pode ser representado como pares ordenados em termos de coordenadas retangulares ou como (r, θ) em termos de coordenadas polares, em que r é a distância do ponto à origem e θ é a direção.

Para o ponto origem O , tem-se que $r = 0$ e não há direção; nesse caso, θ não está definido. Direções no plano podem ser representadas como vetores unitários ou, equivalentemente, como pontos no círculo unitário. Escolhendo uma direção inicial e uma orientação para o círculo unitário, cada ponto pode ser representado com um ângulo θ relacionados como o par ordenado $(\cos \theta, \sin \theta)$ (Morales, 2006).

2.1.1 Estatística descritiva para dados circulares

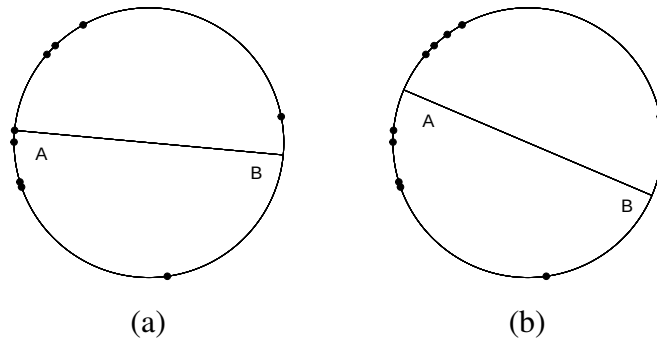
A média circular $\bar{\theta}$ é uma medida de tendência central específica para dados que possuem uma natureza circular, em que o início e o fim do intervalo se conectam. A média circular $\bar{\theta}$ é calculada a partir do cosseno C e seno S do vetor médio:

$$(C, S) = \left(\frac{1}{n} \sum_{i=1}^n \cos \theta_i, \frac{1}{n} \sum_{i=1}^n \sin \theta_i \right). \quad (2.1)$$

Essa fórmula calcula as coordenadas cartesianas associadas a cada observação angular com índice i , resultando no vetor médio (C, S) das observações angulares no plano cartesiano. O ângulo $\bar{\theta}$ em termos de coordenadas polares desse vetor é a média circular (Jammalamadaka; SenGupta, 2001).

Para obter a mediana circular $\tilde{\theta}$, deve-se dividir o círculo em dois semicírculos com igual número de observações de tamanho n a partir do diâmetro formado pelos pontos A e B . Geometricamente, o ponto A , que corresponde à mediana circular $\tilde{\theta}$, está mais próximo em termos de ângulo da maioria das observações, enquanto o ponto B , diametralmente oposto a A , ou seja, a 180° , está mais distante da maioria dos dados. Quando o tamanho da amostra n é ímpar, Figura 2.2 (a), $\tilde{\theta}$ é o ponto que divide as observações pela metade. No caso de n par, Figura 2.2 (b), a divisão dos semicírculos é dada pela média de duas observações adjacentes apropriadas. Ressalta-se que, para uma amostra com n par, usa-se a média circular $\bar{\theta}$ conforme definida na equação (2.1) (Mardia, 1972).

Figura 2.2 – Representação visual da mediana circular com amostras de tamanhos (a) ímpar e (b) par.



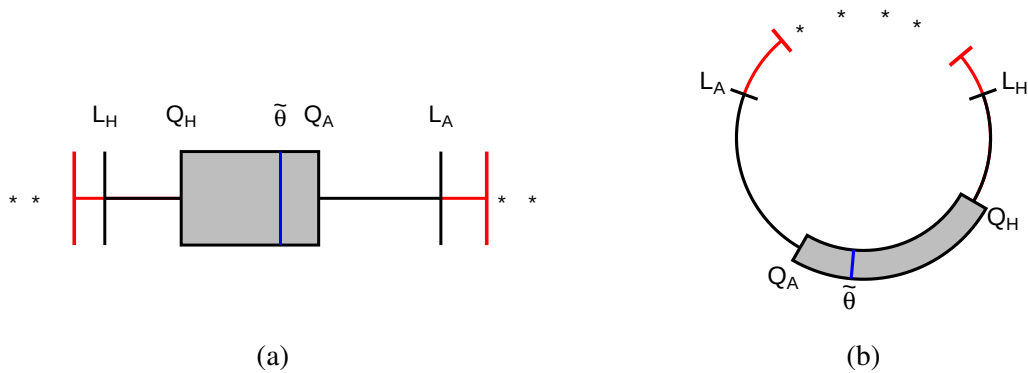
Fonte: Adaptado de MARDIA (1972).

Semelhante à mediana para dados lineares, que possui primeiro, segundo (mediana) e terceiro quartis, Buttarazzi (2018) define os quartis circulares em relação à mediana circular

$\tilde{\theta}$. Os quartis circulares são classificados como o quartil no sentido horário Q_H e o quartil no sentido anti-horário Q_A , visto que os subscritos “H” e “A” representam “sentido horário” e “sentido anti-horário”, respectivamente.

Para construir o boxplot circular (Figura 2.3), primeiramente traça-se um segmento de reta transversal na mediana circular $\tilde{\theta}$. A partir desse ponto, desenha-se a caixa do boxplot, que se estende até os quartis Q_H (horário) e Q_A (anti-horário). Os extremos, destacados em vermelho, são determinados multiplicando a distância interquartílica (comprimento da caixa) por um fator usualmente igual a 1,5. Os limites do boxplot circular, chamados de limite horário L_H e limite anti-horário L_A , equivalem aos “bigodes” de um boxplot linear e são obtidos ajustando os extremos até a observação mais próxima da caixa. Observações além desses limites são classificadas como discrepantes (outliers), de forma análoga aos dados lineares (Buttarazzi; Pandolfo; Porzio, 2018).

Figura 2.3 – Boxplot para o caso (a) linear e (b) circular.



Fonte: Do autor (2025).

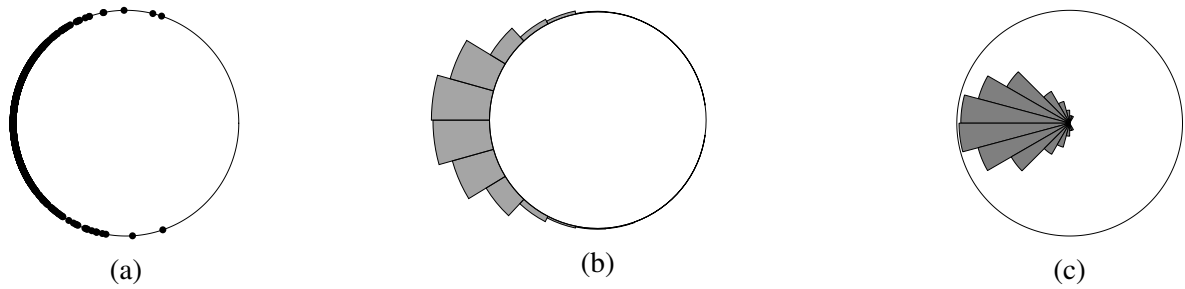
2.1.2 Gráficos para Dados Circulares

A análise visual proporcionada pelo boxplot circular é uma importante ferramenta para identificar a estrutura de dados circulares. No entanto, para explorar padrões mais complexos, outras representações gráficas são igualmente essenciais, como os gráficos de dispersão circular, diagramas de rosa e histogramas circulares.

O diagrama de dispersão circular, Figura 2.4 (a), exibe cada observação como um ponto no círculo unitário, permitindo identificar padrões de agrupamento e densidade angular (Fisher; Lewis; Embleton, 1987). O histograma circular, Figura 2.4 (b) representa a frequência de obser-

vações em intervalos angulares, com barras dispostas ao longo do contorno circular, destacando picos de concentração em distribuições unimodais ou multimodais (Pewsey; Neuhaus; Davis, 2013). Já o diagrama de rosa, uma variação do histograma, apresenta barras com origem no centro do círculo, proporcionais à densidade de dados, Figura 2.4 (c).

Figura 2.4 – Gráficos para dados circulares: (a) diagrama de dispersão circular, (b) histograma circular, (c) diagrama de rosa.



Fonte: Do autor (2025).

2.1.3 Modelos probabilísticos para dados circulares

Entende-se por variável aleatória circular como sendo uma variável aleatória já convertida em ângulos e assumindo valores no intervalo $[0, 2\pi)$. Define-se a função de distribuição F de uma variável aleatória circular Y como sendo uma função F tal que:

$$F(y) = P(0 < Y \leq y), \quad \forall y \in [0, 2\pi),$$

$$F(y + 2\pi) - F(y) = 1, \quad \forall y \in \mathbb{R}.$$

A função de densidade de probabilidade associada, denotada por f , descreve a distribuição de probabilidade de Y ao longo do círculo. Para $-\infty < y_1 < y_2 < \infty$, com $y_2 - y_1 < 2\pi$, a probabilidade de Y pertencer ao arco (y_1, y_2) é dada por:

$$P(y_1 < Y \leq y_2) = F(y_2) - F(y_1) = \int_{y_1}^{y_2} f(y) dy.$$

Embora $F(0) = 0$ e $F(2\pi) = 1$, para as variáveis aleatórias circulares, a função de distribuição no círculo não satisfaz as condições $\lim_{x \rightarrow -\infty} F(0) = 0$ e $\lim_{x \rightarrow \infty} F(2\pi) = 1$, que são satisfeitas por funções de distribuições de variáveis aleatórias lineares (Souza, 1999).

2.1.3.1 Distribuição circular uniforme

Se a probabilidade total está espalhada uniformemente sobre a circunferência, temos a distribuição uniforme circular U com função densidade de probabilidade f da forma:

$$f(y) = \frac{1}{2\pi}, \quad 0 \leq y < 2\pi.$$

Observe que, se $\theta \leq \beta \leq \alpha + 2\pi$, então:

$$P_r(\alpha < Y \leq \beta) = F(\beta) - F(\alpha) = \int_{\alpha}^{\beta} f(y) dy.$$

Logo,

$$P_r(\alpha < Y \leq \beta) = \frac{1}{2\pi} y \Big|_{\alpha}^{\beta} = \frac{\beta - \alpha}{2\pi}.$$

Em uma distribuição circular uniforme, todas as direções têm igual densidade e, por isso, essa distribuição é comumente conhecida como distribuição isotrópica ou aleatória e por isso não tem média bem definida. Ela é muito importante na análise de dados circulares, pois representa a falta de “direção preferencial”. Quando a distribuição não é uniforme, espera-se uma concentração ao redor de uma ou mais direções. Usualmente, tem-se uma concentração maior ao redor de uma única direção e a distribuição é unimodal (Neves, 2015).

2.1.3.2 Distribuição von Mises

Outros modelos probabilísticos são obtidos por intermédio de analogias no círculo com importantes caracterizações univariadas, dando origem a igualmente importantes distribuições de probabilidade para variáveis aleatórias circulares. A distribuição de von Mises vM é tão importante para dados circulares quanto o modelo Gaussiano é para dados lineares em geral (Souza, 1999). Uma variável aleatória circular Y tem distribuição von Mises vM com parâmetros de média circular μ e concentração κ se sua função densidade de probabilidade f for dada por:

$$f(y, \mu, \kappa) = \frac{1}{\sqrt{2\pi I_0(\kappa)}} \exp\{\kappa \cos(y - \mu)\},$$

em que $0 \leq y < 2\pi$, $0 \leq \mu < 2\pi$, $\kappa > 0$ e $I_0(\kappa)$ é a função de Bessel modificada de primeiro tipo e ordem zero, que é dada por:

$$I_0(\kappa) = \sum_{k=-\infty}^{\infty} \frac{1}{(r!)^2} \left(\frac{1}{2}\kappa\right)^{2r}.$$

Na distribuição von Mises, o parâmetro μ representa a direção média ou preferencial, funcionando como uma medida de localização no círculo, análoga à média em uma distribuição normal. Já o parâmetro κ é um parâmetro de concentração, que controla a dispersão das observações em torno de μ . Quando $\kappa = 0$, a distribuição se torna uniforme, indicando ausência de direção preferencial; à medida que κ aumenta, a densidade se concentra mais fortemente em torno de μ , assemelhando-se a uma distribuição normal no limite para κ muito grande (Souza, 1999).

A função de distribuição de Y é dada por:

$$F(y) = \frac{1}{\sqrt{2\pi I_0(\kappa)}} \int_0^y \exp\{\kappa \cos(y - \mu)\} dy.$$

$F(y)$ não admite uma forma fechada (Souza, 1999).

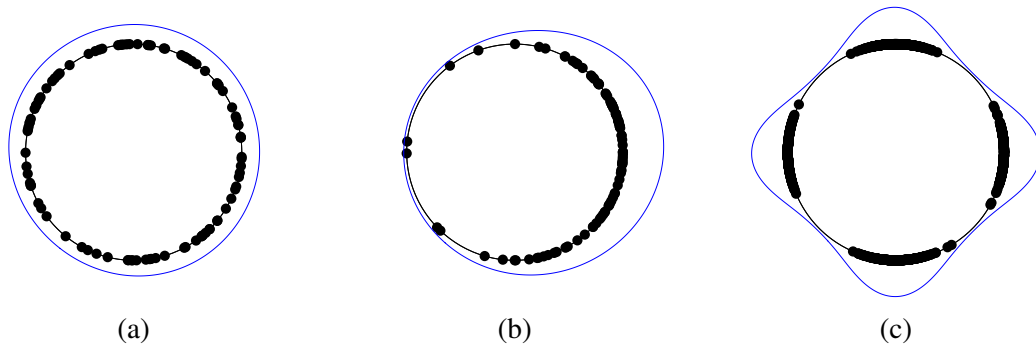
Para avaliar a adequação de dados circulares às distribuições uniforme ou von Mises, utilizou-se os testes de Watson e de Espaçamento de Rao, implementados no pacote `circular` (Lund; Agostinelli; Agostinelli, 2017) do software R (Team, 2025).

2.2 Multimodalidade na análise de dados

Quanto aos diferentes padrões de distribuição dos dados, se não há uma direção média bem definida, diz-se que os dados estão uniformemente distribuídos no círculo, seguindo a distribuição uniforme circular U (Figura 2.5 (a)). No entanto, se os dados tendem a uma única direção média bem definida, a distribuição dos dados é unimodal (Figura 2.5 (b)). Caso contrário, há distribuições multimodais, ou M -modais, em que o número de modas, por exemplo, $M=4$ representa uma distribuição 4-modal (Figura 2.5 (c)).

A multimodalidade em dados lineares e circulares apresenta diferenças fundamentais devido à natureza distinta dos espaços em que os dados são representados, como ilustrado na Figura 2.6. Em distribuições lineares bimodais (Figura 2.6 (a)), a separação entre os dois agru-

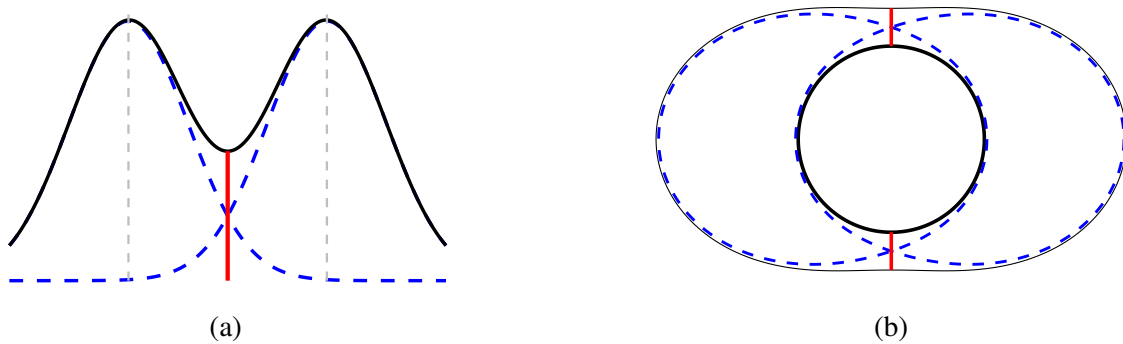
Figura 2.5 – Distribuições no círculo: (a) uniforme, (b) unimodal e (c) 4-modais simétricas.



Fonte: Do autor (2025).

pamentos requer apenas um único ponto de segmentação, já que os valores menores tendem a formar um grupo e os maiores, outro, permitindo uma divisão clara com base em um limiar. Por outro lado, em distribuições circulares bimodais (Figura 2.6 (b)), a ciclicidade do espaço angular complica essa separação, exigindo dois pontos de segmentação. Isso ocorre porque, no círculo, não há noção clara de "extremos" como em uma reta, e os dados podem se agrupar em regiões opostas ou adjacentes, demandando uma abordagem que considere a continuidade angular para identificar corretamente os modos.

Figura 2.6 – Representação de dados bimodais (a) lineares e (b) circulares.



Fonte: Do autor (2025).

Detectar desvios da uniformidade torna-se mais desafiador quanto maior for o valor de M , ou seja, o número de modas. Landler et al. (2018) recomendam que, para distribuições M -modais simétricas, os ângulos originais sejam multiplicados por M , mantendo-se o módulo em relação ao ângulo máximo permitido, de 0 a 2π . Esse procedimento converte a distribuição em unimodal, melhorando assim, a capacidade dos testes estatísticos de detectarem padrões não uniformes.

No entanto, essa abordagem acarreta consequências importantes. Ela é eficaz em casos de simetria perfeita, onde os modos são igualmente espaçados e de mesma forma, uma condição raramente atendida em dados reais, porém útil em simulação de dados. Assim, as discrepâncias entre os agrupamentos, como diferenças em suas dispersões, assimetrias ou posições, podem reduzir a capacidade do teste de distinguir a distribuição transformada de uma uniforme. Além disso, exige a especificação a priori do número de modos (M), e escolhas incorretas podem comprometer a validade dos resultados.

2.2.1 Métodos de segmentação em dados circulares multimodais

Decidir sobre uma possível multimodalidade de um conjunto de dados é um problema importante na análise de dados e modelagem estatística. Deve-se conhecer a origem dos pontos, isto é, se foram gerados por uma distribuição de probabilidade com um único pico ou mais de um pico (Chasani; Likas, 2022).

No caso de distribuições lineares é crucial encontrar estimativas para as modas para compreender conjuntos de dados complexos. Da mesma forma, detectar vales, ou seja, regiões com escassez, ou mesmo ausência de pontos, auxilia na definição de possíveis agrupamentos. Esses dois métodos são essenciais em uma análise de dados multimodais (Chasani; Likas, 2024).

No caso dos dados circulares, tal análise também é útil, no entanto, para realizar as análises estatísticas, um dos procedimentos utilizados é, após identificar a multimodalidade, transformar os dados para gerar uma distribuição unimodal.

Para tal realização, o passo inicial é estabelecer os limites de dados que pertencem a cada grupo que contém a moda. Isso pode ser realizado de maneira empírica, ou por meio de um método probabilístico, utilizando o algoritmo EM, por exemplo. A seguir, uma breve descrição desses dois métodos.

2.2.1.1 Método Empírico

Esse método consiste em uma abordagem em que o início e o fim do ciclo são determinados de maneira arbitrária, não levando em consideração as características dos dados no ciclo. Esses pontos de segmentação geralmente correspondem a convenções cotidianas, como definir

23h59min como o fim de um agrupamento em função do dia do calendário e 00h como o início de outro.

O algoritmo de segmentação pelo Método Empírico pode ser resumido da seguinte forma:

- a) determinar o número de modas M de grupos;
- b) encontrar os M pontos de segmentação θ_{Emp}^* :

$$\theta_{EMP}^* = \theta_I^* + (m - 1) \frac{360^\circ}{M} \quad \text{para } m = 1, 2, \dots, M \quad (2.2)$$

em que θ_I^* é o ponto inicial, que, no Método Empírico, geralmente se adota 0° .

Seguindo esta metodologia, tem-se, por exemplo, para dados que têm influência dos dias da semana, poderiam ser categorizados da seguinte maneira:

- a) domingo: $[0.00^\circ; 51.43^\circ]$;
- b) segunda: $[51.43^\circ; 102.86^\circ]$;
- c) terça: $[102.86^\circ; 154.29^\circ]$;
- d) quarta: $[154.29^\circ; 205.71^\circ]$;
- e) quinta: $[205.71^\circ; 257.14^\circ]$;
- f) sexta: $[257.14^\circ; 308.57^\circ]$;
- g) sábado: $[308.57^\circ; 360.00^\circ]$.

Ou seja, 0° como as 0 horas e 0 minutos do domingo, e cada um dos 7 dias teria o comprimento (duração) de $\frac{360^\circ}{7}$.

2.2.1.2 Método Probabilístico por algoritmo EM

O algoritmo EM foi proposto inicialmente por Dempster et al.(1977) e implementado por Hornik e Grün no pacote de software estatístico R movMF (2021). Em cada iteração do algoritmo EM, há duas etapas chamadas: Etapa de Expectativa (E) e Etapa de Maximização (M). As etapas E e M são repetidas consecutivamente até que os critérios de convergência sejam atingidos (Bruzzone; Prieto, 2002).

O algoritmo EM utiliza a probabilidade P_{nk} de cada ponto de dados x_n pertencer a um agrupamento com índice $k = 1, \dots, K$, a função de densidade de probabilidade f da distribuição von Mises vM , a proporção ω_k de cada grupo, os parâmetros de média angular μ_k , os parâmetros

de concentração κ_k , e a função de verossimilhança $\mathcal{L}(\mu, \kappa, \pi)$. As duas etapas principais do algoritmo EM são descritas da seguinte forma:

- a) expectativa (E): nessa etapa, calcula-se a probabilidade P_{nk} de cada ponto de dados x_n pertencer a cada agrupamento com índice k , dado os parâmetros atuais do modelo. A probabilidade P_{nk} é dada por:

$$P_{nk} = \frac{\omega_k f(x_n | \mu_k, \kappa_k)}{\sum_{j=1}^K \omega_j f(x_n | \mu_j, \kappa_j)} \quad (2.3)$$

em que ω_k é a proporção esperada para o grupo com índice k e $f(x_n | \mu_k, \kappa_k)$ é a função de densidade da distribuição von Mises para o agrupamento k , avaliada no ponto x_n ;

- b) maximização (M): nessa etapa, os parâmetros do modelo são atualizados para maximizar a função de verossimilhança $\mathcal{L}(\mu, \kappa, \omega)$. A função de verossimilhança $\mathcal{L}(\mu, \kappa, \omega)$ é dada por:

$$\mathcal{L}(\mu, \kappa, \omega) = \sum_{n=1}^N \sum_{k=1}^K P_{nk} \log [\omega_k f(x_n | \mu_k, \kappa_k)] \quad (2.4)$$

em que μ_k são os parâmetros de média angular, κ_k são os parâmetros de concentração dos agrupamentos, e ω_k são as proporções de cada grupo com índice k .

A matriz de probabilidade P_{nk} que fornece a probabilidade de cada ponto pertencer a um determinado agrupamento foi obtida utilizando o pacote `movMF` (Hornik; Leisch, 2014) do software R (Team, 2025). Esse pacote implementa modelos de mistura de distribuições von Mises, permitindo a modelagem de dados circulares e esféricos, e será utilizado para calcular as probabilidades de alocação dos pontos de dados aos agrupamentos durante o passo de Expectativa do algoritmo EM.

O modelo de mistura é uma ferramenta popular e amplamente utilizado para agrupamento devido a seus fundamentos probabilísticos e flexibilidade (Bouveyron; Brunet-Saumard, 2014). Nesse método, os dados são atribuídos aos agrupamentos com base na distribuição do componente com maior probabilidade de tê-lo gerado. O procedimento de agrupamento envolve ajustar um modelo de mistura na qual cada ponto é associado ao componente com a maior probabilidade (Chasani; Likas, 2024).

2.2.2 Critérios de Informação AIC e BIC

Para avaliar cada modelo de mistura serão calculados os critérios de informação de Akaike (AIC) e Bayesiano (BIC). No software R, esses critérios são implementados no pacote `stats` (Team, 2025) por meio das funções `AIC` e `BIC`, que calculam os valores com base no logaritmo de máxima verossimilhança, definidas como:

a) critério de informação de Akaike (AIC):

$$\text{AIC} = -2 \cdot \ell(\hat{\theta}) + k \cdot n_{\text{par}}$$

onde $\ell(\hat{\theta})$ é o logaritmo da verossimilhança máxima do modelo ajustado com os parâmetros estimados $\hat{\theta}$, n_{par} é o número de parâmetros do modelo, e k é a penalização por parâmetro, com valor padrão $k = 2$ para o AIC;

b) critério de informação Bayesiano (BIC):

$$\text{BIC} = -2 \cdot \ell(\hat{\theta}) + n_{\text{par}} \cdot \ln(n)$$

onde $\ell(\hat{\theta})$ e n_{par} são como definidos no AIC, e n é o número de observações na amostra.

2.3 Análise de correlação entre variáveis

Explorou-se a correlação entre diferentes tipos de variáveis quantitativas para compreender as relações entre os fatores ambientais e de configuração dos ninhos-armadilhas, para variáveis quantitativas lineares foi calculado o coeficiente de correlação de Pearson, para variáveis lineares e circulares utilizou-se a correlação circular-linear, e para pares de variáveis circulares aplicou-se a correlação circular de Fisher.

A correlação linear-linear (LL) é aplicada quando ambas as variáveis são quantitativas e pertencem ao domínio da reta real. Para avaliar a associação linear entre duas variáveis X e Y , com médias amostrais $\bar{x}$ e $\bar{y}$, o coeficiente de correlação de Pearson r_{xy} é definido como:

$$LL = r(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

no qual x_i e y_i representam os valores observados, e $\bar{x}$ e $\bar{y}$ são as médias amostrais respectivas. Essa medida quantifica o grau de associação linear entre variáveis e, para inferências estatísticas, idealmente requer uma distribuição aproximadamente normal. É amplamente aplicada em estudos com variáveis contínuas no conjunto dos números reais (Asuero; Sayago; González, 2006).

A correlação linear-circular (LC) é empregada quando uma variável é linear, situada na reta real, e a outra é circular, representada como um ângulo em um círculo unitário. Uma medida de correlação entre uma variável linear x e uma variável angular θ é a correlação linear-circular (Mardia, 1976). Considere uma amostra dada por $(x_1, \theta_1), \dots, (x_n, \theta_n)$, na qual x_i é a variável linear e θ_i é circular; a correlação é dada por:

$$LC = r(X, \theta) = \frac{r_{xc}^2 + r_{xs}^2 - 2r_{xc}r_{xs}r_{cs}}{1 - r_{cs}^2},$$

em que:

- a) $r_{xc} = r(x_i, \cos \theta_i)$;
- b) $r_{xs} = r(x_i, \sin \theta_i)$;
- c) $r_{cs} = r(\cos \theta_i, \sin \theta_i)$.

A correlação circular-circular (CC) é utilizada quando ambas as variáveis são angulares, localizadas em um círculo unitário, como a observação angular θ e a direção angular ϕ , com médias direcionais $\bar{\theta}$ e $\bar{\phi}$. Um coeficiente de correlação CC é calculado como:

$$CC = r(\theta, \phi) = \frac{\sum_{i=1}^n [\sin(\theta_i - \bar{\theta}) \sin(\phi_i - \bar{\phi})]}{\sqrt{\sum_{i=1}^n \sin^2(\theta_i - \bar{\theta}) \sum_{i=1}^n \sin^2(\phi_i - \bar{\phi})}},$$

no qual θ_i e ϕ_i são as variáveis circulares, e $\bar{\theta}$ e $\bar{\phi}$ são suas médias direcionais, obtidas por métodos de estatística circular (Jammalamadaka; Sengupta, 2001).

Para a variável qualitativa cor de revestimento, utilizou-se o coeficiente V de Cramer, calculado com base na estatística do teste Qui-Quadrado de Pearson χ^2 , no total de observações n , no número de linhas m da tabela de contingência, e no menor valor entre os graus de liberdade $\min(m-1, n-1)$ (Santos; Carneiro; Oliveira, 2016), dado por:

$$V = \sqrt{\frac{\chi^2}{n \cdot \min(m-1, n-1)}},$$

em que N é o total de observações na amostra, m é o número de linhas da tabela de contingência, n é o número de colunas da tabela de contingência e $\min(m-1, n-1)$ é o menor valor entre os graus de liberdade das linhas ($m-1$) e das colunas ($n-1$). O coeficiente V varia

entre 0 (nenhuma associação) e 1 (associação perfeita), indicando a força da relação entre as variáveis qualitativas analisadas.

2.4 Dados de Sobrevivência

A análise de sobrevivência estuda o tempo de sobrevivência T até a ocorrência de um evento de interesse, frequentemente lidando com dados incompletos devido à censura (Turkson; Ayiah-Mensah; Nimoh, 2021). A variável aleatória T é contínua e não negativa, acompanhada por uma variável indicadora $Y_i(t)$, que determina se o indivíduo i está sob risco no instante t :

$$Y_i(t) = \begin{cases} 0, & \text{se o indivíduo } i \text{ não está sob risco no instante } t; \\ 1, & \text{se o indivíduo } i \text{ está sob risco no instante } t. \end{cases}$$

Essa variável indicadora é registrada como uma informação adicional, assim como possíveis covariáveis. Pode-se definir três funções para descrever os diferentes aspectos apresentados pelos dados de sobrevivência (Tomazella, 2003):

- a) função densidade de probabilidade: a função densidade $f(t) > 0$ para todo t maior que zero é definida como o limite da probabilidade da ocorrência do evento de um indivíduo no intervalo de tempo $[t, t + dt)$, expressa por (Cox; Oakes, 1984):

$$f(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T \leq t + dt)}{dt};$$

- b) função de sobrevivência: a função $S(t)$ descreve a forma distribucional dos tempos de sobrevivência por meio da probabilidade da não ocorrência do evento de interesse até o instante t , sendo denominada função de sobrevivência. É uma função não crescente, com $S(0) = 1$ e $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$, dada por (Lawless, 1982):

$$S(t) = P(T \geq t);$$

- c) função de risco: a função de risco, ou taxa de falha, é definida pelo limite da probabilidade de um indivíduo falhar no intervalo $[t, t + dt)$, dado que não tenha falhado

até o instante t , expressa por (Cox; Oakes, 1984):

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt \mid T \geq t)}{dt} = \frac{f(t)}{S(t)}.$$

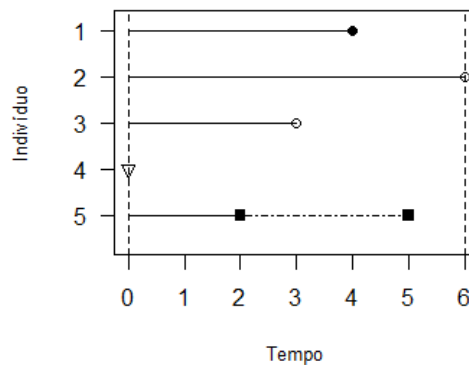
2.4.1 Mecanismo de Censura

A censura ocorre quando o tempo exato do evento não é observado, resultando em informação parcial (Andersen *et al.*, 2021). A Figura 2.7 representa os principais tipos de censura, que são:

- a) censura à direita: o evento não é observado até o final do estudo, seja por término do período de observação, número fixo de eventos ou retirada por causas externas;
- b) censura à esquerda: o evento ocorreu antes do início do estudo, mas o momento exato é desconhecido;
- c) censura intervalar: o evento é conhecido por ter ocorrido em um intervalo $[T_0, T_F]$, no qual T_0 é o último momento sem evento e T_F é o momento em que o evento é verificado.

Figura 2.7 – Representação gráfica de registros de tempos com censura e sem censura.

Legenda: (- -) pontilhado vertical representa o início e fim do estudo; (—) retas horizontais corresponde ao período que o indivíduo esteve sob risco; (●) a ocorrência do evento; (○) censura à direita; (▽) censura à esquerda; (- · -) pontilhado horizontal representa a censura intervalar com (■) sendo os intervalos de início e fim.



Fonte: Do autor (2025).

Na Figura 2.7, somente o indivíduo 1 teve a ocorrência do evento antes do fim do estudo, no tempo $t = 4$. O indivíduo 2 não teve ocorrência do evento durante o prazo estabelecido pelo estudo. Caso o prazo tenha sido determinado por uma unidade de tempo pré-estabelecida,

trata-se de censura à direita do tipo I; se o prazo tiver sido determinado por uma quantidade pré-determinada de falhas, então se trata de censura à direita do tipo II. O indivíduo 3 foi retirado do experimento por um motivo diferente do evento de interesse, configurando uma censura à direita aleatória. O indivíduo 4 já havia experimentado o evento ao iniciar o estudo, representado pela censura à esquerda no tempo $t = 0$. O indivíduo 5 não teve a ocorrência de seu evento marcada no exato momento da falha, mas sabe-se que ocorreu entre os tempos $T_0 = 2$ e $T_F = 5$. Também é possível afirmar que o indivíduo 5 não foi avaliado nos tempos $t = 3$ e $t = 4$, uma vez que o último tempo em que se sabe que não houve ocorrência da falha foi em $T_0 = 2$, e o instante em que foi verificado que a falha já havia ocorrido foi em $T_F = 5$.

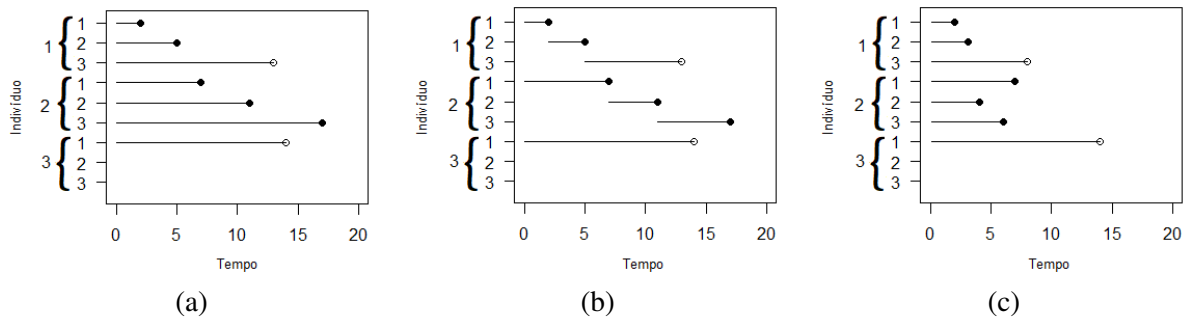
2.4.2 Eventos Recorrentes

Eventos recorrentes referem-se à ocorrência repetida do mesmo evento em uma unidade experimental, como múltiplas capturas de enxames em um mesmo ninho-armadilha (Chiou *et al.*, 2023). Esses eventos requerem modelos que considerem a estrutura temporal das ocorrências, permitindo acomodar tanto a independência quanto a dependência entre os eventos, conforme capturado pelos registros de tempo t_{ij} para o i -ésimo objeto e j -ésima ocorrência. Os intervalos sob risco podem ser registrados por diferentes métodos, como ilustrado na Figura ??, que apresenta representações gráficas dos métodos Total Time (TT), Counting Process (CP) e Gap Time (GT). Seguindo (Kelly; Lim, 2000), denotamos o tempo t_{ij} como o tempo da j -ésima ocorrência do evento para o i -ésimo objeto de estudo, onde i indexa os ninhos-armadilhas e j indexa os eventos recorrentes.

- a) total Time (TT): assume independência entre as ocorrências j , com intervalo de risco $[0, t_{ij}]$, onde t_{ij} é o tempo da j -ésima ocorrência em i ;
- b) counting Process (CP): considera dependência entre os eventos, assumindo que o tempo inicial $t_{i,j+1}$ ocorre somente após o tempo final $t_{i,j}$;
- c) gap Time (GT): reinicia o tempo após cada ocorrência, com intervalo de risco $[0, g_{ij}]$, onde g_{ij} é o intervalo entre ocorrências consecutivas do sujeito i , dado por $g_{ij} = t_{ij} - t_{i,j-1}$, modelando a dependência temporal.

A Figura 2.8 (a), ilustra-se o método Total Time (TT), em que o primeiro indivíduo apresenta o primeiro evento no intervalo $(0, 2)$, o segundo no intervalo $(0, 5)$ e uma censura em

Figura 2.8 – Representação gráfica de registros dos tempos: (a) Total Time, (b) Counting Process e (c) Gap Time.



Fonte: Adaptado de Kelly; Lim (2000).

(0, 13). O segundo indivíduo registra o primeiro evento em (0, 7), o segundo em (0, 11) e o terceiro em (0, 17). O terceiro indivíduo apresenta apenas uma censura no intervalo (0, 14).

No método Counting Process (CP) (Figura 2.8 (b)) o final dos intervalos é idêntico ao método TT, porém o intervalo sob risco de um evento é registrado a partir da ocorrência do evento anterior. Já no método Gap Time (GT) (Figura 2.8 (c)) os intervalos tem o mesmo comprimento do método GT, porém todos iniciando em 0.

2.4.3 Modelos na Análise de Sobrevivência

Modelos de regressão na análise de sobrevivência são ferramentas fundamentais para avaliar como covariáveis influenciam o tempo até a ocorrência de um evento de interesse, sendo divididos em três categorias principais: não paramétricos, paramétricos e semiparamétricos (Selingerova; Katina; Horova, 2021). Esses modelos permitem capturar relações entre variáveis explicativas e a probabilidade de sobrevivência, adaptando-se a diferentes contextos e características dos dados.

Os modelos não paramétricos oferecem uma abordagem sem suposições sobre a distribuição do tempo de sobrevivência. Um exemplo clássico é o estimador de Kaplan-Meier, que calcula a função de sobrevivência $S(t)$ diretamente a partir dos tempos de evento e censura observados (Kaplan; Meier, 1958). Esse não consegue estimar diretamente funções de sobrevivência usando dados censurados por intervalo. Para resolver isso, o método de Turnbull (Turnbull, 1976) foi desenvolvido como uma abordagem não paramétrica para estimar funções de sobrevivência diretamente a partir de dados censurados por intervalo (Iijima; Momozaki; Ando, 2025).

O modelos paramétricos são aqueles em que $S(t)$ é determinada por parâmetros com distribuições de probabilidades conhecidas. As vantagens de se presumir que os dados são provenientes de uma determinada distribuição de probabilidade são as inferências que podem ser feitas sobre os parâmetros da distribuição (Moncharmon *et al.*, 2003). Para estimar os parâmetros, na presença de censura, há um consenso na utilização do método de máxima verossimilhança (Silva, 2016),

2.4.4 Modelo Semiparamétrico de Cox

O modelo de Cox (Cox, 1972), um modelo semiparamétrico, assume que o risco $\lambda(t)$ é proporcional entre indivíduos, expresso como:

$$\lambda(t_i | x_i) = \lambda_0(t_i) \exp(x_i' \beta),$$

em que $\lambda_0(t)$ é o risco basal, uma função não paramétrica e não negativa, x_i representa as covariáveis do sujeito i , e β é o coeficiente das covariáveis, parte paramétrica. A função de verossimilhança parcial é usada para estimar β , sem necessidade de modelar $\lambda_0(t)$ (Chung *et al.*, 2018).

2.4.5 Extensão do Modelo de Cox para dados recorrentes: Modelo de Andersen-Gill

A diferença entre o modelo de Cox clássico, descrito anteriormente, e os modelos para eventos recorrentes está na definição do conjunto de indivíduos em risco. No modelo clássico, o indivíduo deixa de estar em risco após experimentar o evento. Para eventos recorrentes, porém, o indivíduo permanece em risco após cada ocorrência do evento de interesse (Tomazella, 2003).

Diversas extensões de modelos de sobrevivência baseadas na abordagem de riscos proporcionais de Cox foram propostas para lidar com dados de eventos recorrentes (Kelly; Lim, 2000), sendo uma delas o modelo de Andersen-Gill (AG) (Andersen; Gill, 1982). O modelo AG considera que todas as ocorrências de um sujeito contribuem para o conjunto de risco, independentemente do número de eventos anteriores, utilizando uma estrutura não estratificada.

A função de risco é expressa como:

$$\lambda(t_{ij} | x_i) = \lambda_0(t_{ij}) \exp(x_i' \beta),$$

onde t_{ij} é o tempo da j -ésima ocorrência para o i -ésimo sujeito, conforme definido anteriormente, x_i são as covariáveis do sujeito i , e $\lambda_0(t_{ij})$ é o risco basal comum, não estratificado, ou seja, o mesmo para todas as ocorrências e sujeitos em cada tempo t_{ij} (Andersen; Gill, 1982; Kelly; Lim, 2000). Essa abordagem agrupa todas as recorrências de cada sujeito no conjunto de risco, permitindo a modelagem da dependência temporal entre eventos por meio de covariáveis e ajustes robustos de variância (Carvalho *et al.*, 2011).

A escolha do modelo AG, cujo o registro de tempo é o *counting process*, para este estudo justifica-se por sua adequação à análise de capturas recorrentes de enxames em ninhos-armadilhas, em que o ninho $j + 1$ só é colocado após a captura de enxame no ninho j .

SEGUNDA PARTE

Artigo 1 - Elaborado conforme a norma NBR 6022	
Título do artigo:	Método de segmentação por boxplot circular
Autores:	Luiz Fernando Silva Resende Carla Regina Guimarães Brighenti



ARTIGO 1 – METODO DE SEGMENTAÇÃO POR BOXPLOT CIRCULAR CIRCULAR BOXPLOT SEGMENTATION METHOD¹

Luiz Fernando Silva Resende

<https://orcid.org/0000-0003-4697-6503>

Carla Regina Guimarães Brighenti

<https://orcid.org/0000-0002-7822-3744>

Resumo: Este estudo propõe a utilização do boxplot circular como técnica para identificação de limites em distribuições circulares temporais com multimodalidade. O método utiliza os quartis circulares para estabelecer segmentações entre ciclos, aplicável a casos multimodais simétricos, transformando dados multimodais em unimodais via multiplicação pelo número de modas e calculando pontos de segmentação baseados na amplitude do boxplot. Através de simulações, compara-se o método proposto com técnicas empíricas e probabilísticas (algoritmo EM), avaliando a eficiência na segmentação de dados multimodais. Os resultados indicam que o método baseado no boxplot circular é mais eficiente que o método empírico, e tão bom quanto o probabilístico, especialmente para altas concentrações circular. O método de segmentação por boxplot circular é simples, exigindo apenas a determinação de quartis circulares, em contraste com a complexidade do algoritmo EM do método probabilístico.

Palavras-chave: dados circulares; multimodalidade; estatísticas circulares; distribuição von Mises; simulação.

Abstract: This study proposes the use of the circular boxplot as a technique for identifying boundaries in temporal circular distributions with multimodality. The method uses circular quartiles to establish segmentations between cycles, applicable to symmetric multimodal cases, transforming multimodal data into unimodal via multiplication by the number of modes and calculating segmentation points based on the boxplot amplitude. Through simulations, the proposed method is compared with empirical and probabilistic techniques (EM algorithm), evaluating the efficiency in segmenting multimodal data. The results indicate that the method based on the circular boxplot is more efficient than the empirical method, and as good as the probabilistic one, especially for high concentrations. The circular boxplot segmentation method is simple, requiring only the determination of circular quartiles, in contrast to the complexity of the EM algorithm in the probabilistic method.

Keywords: circular data; multimodality; circular statistics; von Mises distribution; simulation.

¹ Artigo científico apresentado à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária.

1 INTRODUÇÃO

Dados circulares podem ser registrados diretamente como ângulos ou derivados da conversão de dados cíclicos em medidas angulares, mediante o redimensionamento para uma circunferência de comprimento 2π , correspondente a um círculo unitário de raio 1 (Wang, 2013).

Katiyar e Pandey (2013) revisam modelos de radiação solar global em que dados circulares, como latitude e longitude de um local, ângulo de inclinação de painéis solares, registrado como um valor angular em relação ao horizonte para otimizar a captação de radiação, e o azimute, que indica a direção em relação ao norte geográfico, surgem como parâmetros essenciais.

Por outro lado, devido à inclinação da Terra, essa radiação se diferencia nos locais registrados nas diferentes épocas do ano, o que leva à conversão dos dias do ano em ângulos para capturar a variação sazonal, mapeando o ciclo anual do tempo linear para uma representação angular que reflete a posição cambiante do Sol ao longo das estações. O estudo de Katiyar e Pandey (2013) também aborda a radiação solar ao longo do horário do dia, transformando o tempo diário em ângulos que preservam a periodicidade das horas de luz solar, permitindo análises da incidência de energia em intervalos cíclicos como o nascer e o pôr do sol.

Uma importante ferramenta para estudar dados circulares temporais com maior precisão é o boxplot circular, cujo desenvolvimento é relativamente recente. Inicialmente adaptado da técnica linear por Anderson (1993), e mais recentemente refinado por Buttarazzi et al. (2018), que propuseram uma nova abordagem implementada no pacote `bpDir` do software R (Buttarazzi, 2022; Team, 2025). Essa ferramenta facilita a visualização de quartis e outliers em dados circulares, sendo particularmente útil para distribuições unimodais, mas com potencial de extensão para cenários multimodais, como demonstrado em adaptações para dados lineares por Qarmalah et al. (2016), que introduziram o conceito de *k-boxplots* para capturar múltiplas modas. No entanto, assim como em dados lineares, os dados circulares podem apresentar mais de uma moda, conforme destacado por Chakraborty (2014), onde observações reais frequentemente exibem multimodalidade com vários picos, demandando métodos que vão além da unimodalidade assumida em distribuições como a von Mises (Mises, 1918; Mardia; Jupp, 2000).

A distribuição von Mises, uma generalização da normal para dados circulares, é amplamente utilizada para modelar variáveis unimodais agrupadas em torno de uma média circular μ com concentração κ , e um caso particular com $\kappa = 0$ resulta na distribuição uniforme cir-

cular (Donner; Goldenshluger, 2024). Contudo, para dados multimodais do mundo real, onde observações se agrupam em múltiplas direções, assume-se a necessidade de modelos de mistura, como misturas de von Mises, que segmentam os dados em clusters distintos (Marković; Petrović, 2012; Banerjee *et al.*, 2005). Algoritmos de agrupamento adaptados, como o deslocamento médio baseado em dados circulares de Chang-Chien *et al.* (2012) ou o agrupamento fuzzy de Kesemen *et al.* (2016), e abordagens bayesianas de Rodríguez *et al.* (2020), melhoram a detecção automática de clusters, frequentemente ajustados via algoritmo de Expectativa-Maximização (EM) (Bouveyron; Brunet-Saumard, 2014; Chasani; Likas, 2024). Apesar da eficácia probabilística desses métodos, sua complexidade contrasta com abordagens empíricas mais intuitivas, que definem limites arbitrários (como inícios de ciclos anuais em janeiro), mas falham em capturar estruturas subjacentes (Banerjee *et al.*, 2005).

A complexidade das técnicas probabilísticas de agrupamento para dados circulares pode dificultar sua aplicação, levando ao uso de métodos empíricos determinísticos, que, embora simples, ignoram características cíclicas. Ferramentas gráficas como histogramas circulares e linhas de densidade auxiliam na visualização de multimodalidade, mas o boxplot circular, projetado para unimodalidade, requer adaptações para contextos multimodais (Buttarazzi; Pandolfo; Porzio, 2018).

Diante do exposto este estudo tem como objetivo propor a utilização do boxplot circular como técnica, para identificação de limites das distribuições circulares temporais com presença de multimodalidade, bem como compará-la com a técnica empírica e a probabilística de agrupamento através de simulação, visando facilitar a interpretação das variações temporais.

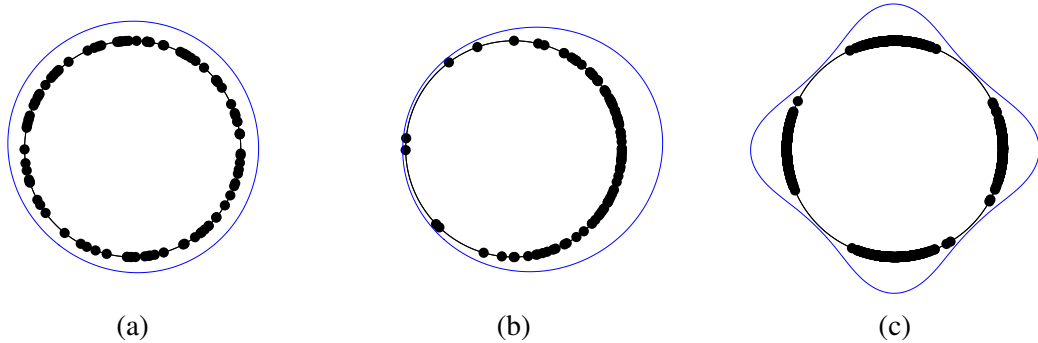
2 REFERENCIAL TEÓRICO

Inicialmente, serão expostas as principais definições necessárias para apresentar e esclarecer o método proposto.

2.1 Multimodalidade na análise de dados

Existem diferentes padrões de distribuição dos dados, se não há uma direção média bem definida, diz-se que os dados estão uniformemente distribuídos no círculo, seguindo a distribuição uniforme circular U (Figura 2.1 (a)). No entanto, se os dados tendem a uma única direção média bem definida, a distribuição dos dados é unimodal (Figura 2.1 (b)). Caso contrário, há distribuições multimodais, ou M -modais, em que o número de modas, por exemplo, $M=4$ representa uma distribuição 4-modal (Figura 2.1 (c)).

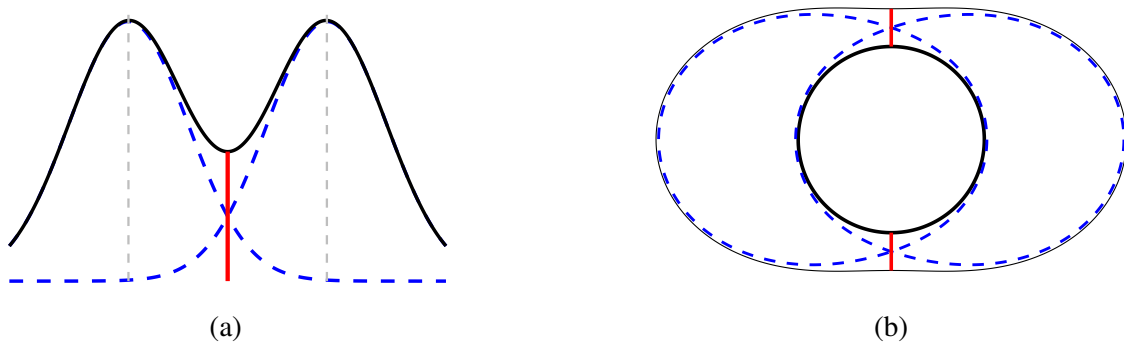
Figura 2.1 – Distribuições no círculo: (a) uniforme, (b) unimodal e (c) 4-modais simétricas.



Fonte: Do autor (2025).

A multimodalidade em dados lineares e circulares apresenta diferenças fundamentais devido à natureza distinta dos espaços em que os dados são representados, como ilustrado na Figura 2.2. Em distribuições lineares bimodais (Figura 2.2 (a)), a separação entre os dois agrupamentos requer apenas um único ponto de segmentação, já que os valores menores tendem a formar um grupo e os maiores, outro, permitindo uma divisão clara com base em um limiar. Por outro lado, em distribuições circulares bimodais (Figura 2.2 (b)), a ciclicidade do espaço angular complica essa separação, exigindo dois pontos de segmentação. Isso ocorre porque, no círculo, não há noção clara de "extremos" como em uma reta, e os dados podem se agrupar em regiões opostas ou adjacentes, demandando uma abordagem que considere a continuidade angular para identificar corretamente os modos.

Figura 2.2 – Representação de dados bimodais (a) lineares e (b) circulares.



Fonte: Do autor (2025).

Detectar desvios da uniformidade torna-se mais desafiador quanto maior for o valor de M , ou seja, o número de modas. Landler et al. (2018) recomendam que, para distribuições M -modais simétricas, os ângulos originais sejam multiplicados por M , mantendo-se o módulo em relação ao ângulo máximo permitido, de 0 a 2π . Esse procedimento converte a distribuição em unimodal, melhorando assim, a capacidade dos testes estatísticos de detectarem padrões não uniformes.

No entanto, essa abordagem acarreta consequências importantes. Ela é eficaz em casos de simetria perfeita, onde os modos são igualmente espaçados e de mesma forma, uma condição raramente atendida em dados reais, porém útil em simulação de dados. Assim, as discrepâncias entre os agrupamentos, como diferenças em suas dispersões, assimetrias ou posições, podem reduzir a capacidade do teste de distinguir a distribuição transformada de uma uniforme. Além disso, exige a especificação a priori do número de modos (M), e escolhas incorretas podem comprometer a validade dos resultados.

2.2 Métodos de segmentação em dados circulares multimodais

Decidir sobre uma possível multimodalidade de um conjunto de dados é um problema importante na análise de dados e modelagem estatística. Deve-se conhecer a origem dos pontos, isto é, se foram gerados por uma distribuição de probabilidade com um único pico ou mais de um pico (Chasani; Likas, 2022).

No caso de distribuições lineares é crucial encontrar estimativas para as modas para compreender conjuntos de dados complexos. Da mesma forma, detectar vales, ou seja, regiões com

escassez, ou mesmo ausência de pontos, auxilia na definição de possíveis agrupamentos. Esses dois métodos são essenciais em uma análise de dados multimodais (Chasani; Likas, 2024).

No caso dos dados circulares, tal análise também é útil, no entanto, para realizar as análises estatísticas, um dos procedimentos utilizados é, após identificar a multimodalidade, transformar os dados para gerar uma distribuição unimodal.

Para tal realização, o passo inicial é estabelecer os limites de dados que pertencem a cada grupo que contém a moda. Isso pode ser realizado de maneira empírica, ou por meio de um método probabilístico, utilizando o algoritmo EM, por exemplo. A seguir, uma breve descrição desses dois métodos.

2.2.1 Método Empírico

Esse método consiste em uma abordagem em que o início e o fim do ciclo são determinados de maneira arbitrária, não levando em consideração as características dos dados no ciclo. Esses pontos de segmentação geralmente correspondem a convenções cotidianas, como definir 23h59min como o fim de um agrupamento em função do dia do calendário e 00h como o início de outro.

O algoritmo de segmentação pelo Método Empírico pode ser resumido da seguinte forma:

- a) determinar o número de modas M de grupos;
- b) encontrar os M pontos de segmentação θ_{Emp}^* :

$$\theta_{EMP}^* = \theta_I^* + (m - 1) \frac{360^\circ}{M} \quad \text{para } m = 1, 2, \dots, M \quad (2.1)$$

em que θ_I^* é o ponto inicial, que, no Método Empírico, geralmente se adota 0° .

Seguindo esta metodologia, tem-se, por exemplo, para dados que têm influência dos dias da semana, poderiam ser categorizados da seguinte maneira:

- a) domingo: $[0.00^\circ; 51.43^\circ]$;
- b) segunda: $[51.43^\circ; 102.86^\circ]$;
- c) terça: $[102.86^\circ; 154.29^\circ]$;
- d) quarta: $[154.29^\circ; 205.71^\circ]$;
- e) quinta: $[205.71^\circ; 257.14^\circ]$;

- f) sexta: $[257.14^\circ, 308.57^\circ[$;
- g) sábado: $[308.57^\circ, 360.00^\circ[$.

Ou seja, 0° como as 0 horas e 0 minutos do domingo, e cada um dos 7 dias teria o comprimento (duração) de $\frac{360^\circ}{7}$.

2.2.2 Método Probabilístico por algoritmo EM

O algoritmo EM foi proposto inicialmente por Dempster et al.(1977) e implementado por Hornik e Grün no pacote de software estatístico R movMF (2021). Em cada iteração do algoritmo EM, há duas etapas chamadas: Etapa de Expectativa (E) e Etapa de Maximização (M). As etapas E e M são repetidas consecutivamente até que os critérios de convergência sejam atingidos (Bruzzone; Prieto, 2002).

O algoritmo EM utiliza a probabilidade P_{nk} de cada ponto de dados x_n pertencer a um agrupamento com índice $k = 1, \dots, K$, a função de densidade de probabilidade f da distribuição von Mises vM , a proporção ω_k de cada grupo, os parâmetros de média angular μ_k , os parâmetros de concentração κ_k , e a função de verossimilhança $\mathcal{L}(\mu, \kappa, \pi)$. As duas etapas principais do algoritmo EM são descritas da seguinte forma:

- a) expectativa (E): nessa etapa, calcula-se a probabilidade P_{nk} de cada ponto de dados x_n pertencer a cada agrupamento com índice k , dado os parâmetros atuais do modelo. A probabilidade P_{nk} é dada por:

$$P_{nk} = \frac{\omega_k f(x_n | \mu_k, \kappa_k)}{\sum_{j=1}^K \omega_j f(x_n | \mu_j, \kappa_j)} \quad (2.2)$$

em que ω_k é a proporção esperada para o grupo com índice k e $f(x_n | \mu_k, \kappa_k)$ é a função de densidade da distribuição von Mises para o agrupamento k , avaliada no ponto x_n ;

- b) maximização (M): nessa etapa, os parâmetros do modelo são atualizados para maximizar a função de verossimilhança $\mathcal{L}(\mu, \kappa, \omega)$. A função de verossimilhança $\mathcal{L}(\mu, \kappa, \omega)$ é dada por:

$$\mathcal{L}(\mu, \kappa, \omega) = \sum_{n=1}^N \sum_{k=1}^K P_{nk} \log [\omega_k f(x_n | \mu_k, \kappa_k)] \quad (2.3)$$

em que μ_k são os parâmetros de média angular, κ_k são os parâmetros de concentração dos agrupamentos, e ω_k são as proporções de cada grupo com índice k .

A matriz de probabilidade P_{nk} que fornece a probabilidade de cada ponto pertencer a um determinado agrupamento foi obtida utilizando o pacote `movMF` (Hornik; Leisch, 2014) do software R (Team, 2025). Esse pacote implementa modelos de mistura de distribuições von Mises, permitindo a modelagem de dados circulares e esféricos, e será utilizado para calcular as probabilidades de alocação dos pontos de dados aos agrupamentos durante o passo de Expectativa do algoritmo EM.

O modelo de mistura é uma ferramenta popular e amplamente utilizado para agrupamento devido a seus fundamentos probabilísticos e flexibilidade (Bouveyron; Brunet-Saumard, 2014). Nesse método, os dados são atribuídos aos agrupamentos com base na distribuição do componente com maior probabilidade de tê-lo gerado. O procedimento de agrupamento envolve ajustar um modelo de mistura na qual cada ponto é associado ao componente com a maior probabilidade (Chasani; Likas, 2024).

2.2.3 Critérios de Informação AIC e BIC

Para avaliar cada modelo de mistura serão calculados os critérios de informação de Akaike (AIC) e Bayesiano (BIC). No software R, esses critérios são implementados no pacote `stats` (Team, 2025) por meio das funções `AIC` e `BIC`, que calculam os valores com base no logaritmo de máxima verossimilhança, definidas como:

- a) critério de informação de Akaike (AIC):

$$\text{AIC} = -2 \cdot \ell(\hat{\theta}) + k \cdot n_{\text{par}}$$

onde $\ell(\hat{\theta})$ é o logaritmo da verossimilhança máxima do modelo ajustado com os parâmetros estimados $\hat{\theta}$, n_{par} é o número de parâmetros do modelo, e k é a penalização por parâmetro, com valor padrão $k = 2$ para o AIC;

- b) critério de informação Bayesiano (BIC):

$$\text{BIC} = -2 \cdot \ell(\hat{\theta}) + n_{\text{par}} \cdot \ln(n)$$

onde $\ell(\hat{\theta})$ e n_{par} são como definidos no AIC, e n é o número de observações na amostra.

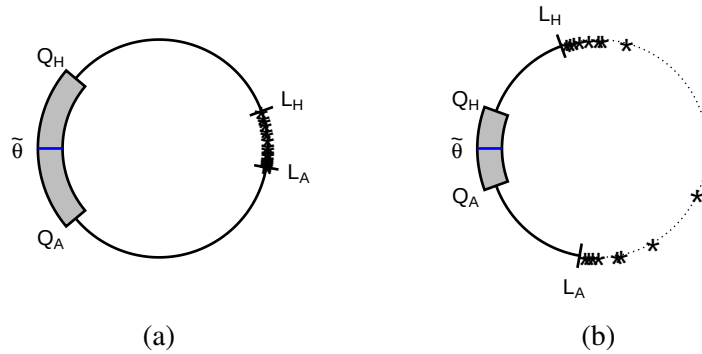
Dito isso, neste trabalho propõe-se um novo método para estabelecer os limites de segmentação de dados circulares utilizando o boxplot circular. Diferentemente do método probabilístico por algoritmo EM, que exige um conhecimento teórico aprofundado em estatística e modelagem probabilística, além de ajustes iterativos complexos para convergência, o boxplot circular é mais simples e direto, aproveitando visualizações gráficas familiares aos pesquisadores. Este método utiliza medidas de tendência central e dispersão adaptadas ao contexto circular, permitindo a detecção de modas e vales. Tal método será descrito a seguir.

3 PROPOSTA DO MÉTODO DE SEGMENTAÇÃO POR BOXPLOT CIRCULAR

O método proposto utiliza os quartis circulares, conforme definidos por Buttarazzi (2018), em dados multimodais, para estabelecer uma segmentação entre os ciclos, as quais pertencem cada uma das modas. Essa metodologia é aplicável a casos multimodais simétricos. Tal simetria é comum em dados circulares temporais, já que seus ciclos possuem o mesmo comprimento (duração), como dias da semana e fases da lua, ou comprimentos bem parecidos como os meses e as estações do ano.

Em situações de dados com M -modas simétricos, a primeira etapa consiste em construir o boxplot supondo uma distribuição unimodal, que é obtida a partir dos dados multimodais. Conforme discutido por Landler e colaboradores (2018), a multiplicação por M será necessária para que o cálculo do boxplot circular seja feito em dados unimodais.

Figura 3.1 – Boxplots circulares de conjuntos de dados gerados a partir de distribuições von Mises: (a) $\kappa = 2$, (b) $\kappa = 5$.



Fonte: Do autor (2025).

Nos casos em que a amplitude do boxplot circular (ABC) é maior que π , ou seja, o arco determinado pelos limites do boxplot como mostra a Figura 3.1 (a), o ponto de segmentação dos ciclos $\theta_{L_i}^*$ será obtido a partir da média circular $\bar{\theta}$ dos limites L_H e L_A , logo $\theta_{L_i}^* = \bar{\theta}_{L_i}$. Na Figura 3.1 (b), como busca-se uma divisão dos ciclos, em uma região da circunferência com menor concentração, o ponto para a segmentação do ciclo será o lado oposto ao da média circular $\bar{\theta}$, portanto, $\theta_{L_i}^* = \bar{\theta}_{L_i} + 180^\circ$.

Após encontrar o ponto de segmentação $\theta_{L_i}^*$ nos dados unimodais, nos dados M -modais os pontos de segmentação θ_m^* serão os pontos proporcionais a $\theta_{L_i}^*$, ou seja:

$$\theta_m^* = \frac{\theta_{L_i}^*}{M} + (m-1) \frac{360^\circ}{M} \quad \text{para } m = 1, 2, \dots, M \quad (3.1)$$

Assim o método proposto trata, basicamente, de formular uma estratégia para alterar o ponto inicial θ_I^* da equação (2.1) no Método Empírico. O algoritmo do Método Boxplot Circular pode ser resumido da seguinte forma:

- a) determinar o número M de grupos;
- b) transformação de dados multimodais para unimodais para dados com M -modas simétricos, conforme Landler et al. (2018):

- multiplique os ângulos dos dados por M para transformar a distribuição multimodal em unimodal;

- c) construção do boxplot circular: determine os limites do boxplot:

- L_A (limite anti-horário): tipicamente, $Q_A - 1.5 \times IQR$;

- L_H (limite horário): tipicamente, $Q_H + 1.5 \times IQR$, na qual IQR é a amplitude interquartílica ($Q_H - Q_A$);

- d) cálculo do ponto de segmentação $\theta_{L_i}^*$ para dados unimodais:

- caso $ABC \geq \pi$: calcule $\theta_{L_i}^*$ como a média circular dos limites:

$$\theta_{L_i}^* = \bar{\theta}_{L_i} = \frac{L_A + L_H}{2};$$

- caso $ABC < \pi$: calcule $\theta_{L_i}^*$ no lado oposto da média circular para capturar a região de menor concentração:

$$\theta_{L_i}^* = \bar{\theta}_{L_i} + 180^\circ;$$

- e) determinação dos pontos de segmentação para dados multimodais: para cada m em uma distribuição M -modal:

- calcule os pontos de segmentação usando:

$$\theta_m^* = \frac{\theta_{L_i}^*}{M} + (m-1) \frac{360^\circ}{M} \quad \text{para } m = 1, 2, \dots, M.$$

4 METODOLOGIA

O desenvolvimento deste trabalho foi dividido em duas etapas principais: a primeira etapa envolveu a simulação da multimodalidade em dados circulares, enquanto a segunda consistiu na verificação da ocorrência de multimodalidade em um conjunto de dados reais sobre horários de acidentes de trânsito. Em ambas as etapas, foram utilizados os dois métodos para a identificação de ciclos, citados na subseção 2.2 e o proposto neste trabalho, denominado Método de Segmentação por Boxplot Circular descrito na seção 3.

4.1 Estudo por simulação da multimodalidade em dados circulares

A simulação foi realizada em três etapas distintas considerando a distribuição von Mises com diferentes parâmetros. Na primeira etapa utilizaram dois diferentes parâmetros de média circular μ e não houve variação do parâmetro de concentração, sendo utilizado apenas o método empírico de segmentação. No segundo caso fez-se, um estudo de simulação com misturas de von Mises para analisar a eficácia dos três métodos de segmentação a serem estudados: (i) Método Empírico, (ii) Método Probabilístico por algoritmo EM e (iii) Método por Boxplot Circular.

4.1.1 Uso do Método Empírico em dados provenientes da distribuição von Mises

No estudo por simulação, foram utilizados conjuntos de dados com índice j para identificar cada conjunto. Foram simulados quatro conjuntos de dados com n_j valores para θ , sendo $n_1 = 5000$, $n_2 = 2000$, $n_3 = 1000$ e $n_4 = 500$, seguindo uma distribuição von Mises. Utilizou-se a função `rvmises` do pacote `circular` no R (2017), com parâmetro $\kappa = 4$ e média circular dada por:

Caso 1: $\mu = 0$.

Caso 2: $\mu = \pi$.

Esses dados foram transformados para que cada um dos quatro conjuntos de dados originais fossem ajustados para ocupar um quarto de círculo. A variável θ foi transformada em θ' usando a equação:

$$\theta' = \frac{\theta}{4} + (j-1) \cdot \frac{\pi}{2} \quad (4.1)$$

em que j representa o índice do conjunto de dados. Essa transformação permite que cada conjunto de dados gerado com os mesmos parâmetros seja distribuído pelos quatro quadrantes, resultando em um conjunto de dados 4-modal. Com base nesses dados θ' de tamanho $n = 8500$, foi realizado o agrupamento de maneira empírica, dividindo os dados em quadrantes. Assim, os grupos de dados $G_{\mu;j}^{Emp}$ foram determinados pelos seguintes critérios:

- a) $G_{\mu;1}^{Emp}$: θ' entre $[0, \frac{\pi}{2}[$;
- b) $G_{\mu;2}^{Emp}$: θ' entre $[\frac{\pi}{2}, \pi[$;
- c) $G_{\mu;3}^{Emp}$: θ' entre $[\pi, \frac{3\pi}{2}[$;
- d) $G_{\mu;4}^{Emp}$: θ' entre $[\frac{3\pi}{2}, 2\pi[$.

Realizou-se a análise de seis maneiras nesses dados simulados organizados do seguinte modo: $G_{\mu;1}^{Emp}$, $G_{\mu;2}^{Emp}$, $G_{\mu;3}^{Emp}$ e $G_{\mu;4}^{Emp}$ e outros dois considerando a totalidade dos dados ($n = 8500$), diferenciando-se pelo posicionamento dos dados. O primeiro conjunto é o unimodal, que se refere à junção dos quatro conjuntos θ com parâmetros μ e κ , e para o segundo conjunto de tamanho 8500, considerou-se θ' , portanto um conjunto de dados 4-modal.

Estatísticas descritivas circulares foram calculadas para cada uma das 6 configurações dos dados usando a função `mle.vonmises` do pacote `circular` (Lund; Agostinelli; Agostinelli, 2017).

Os dados foram testados, para uniformidade e distribuição von Mises, utilizando o teste de espaçamento de Rao, teste de Watson, para a distribuição uniforme e teste de Watson para a distribuição von Mises, respectivamente, pelas funções `rao.test` e `watson.test` do pacote `circular` (Lund; Agostinelli; Agostinelli, 2017), ao nível de significância de 5%. A análise da primeira simulação foi repetida para a segunda, e os resultados foram comparados.

4.1.2 Estudo da multimodalidade em misturas de distribuição von Mises

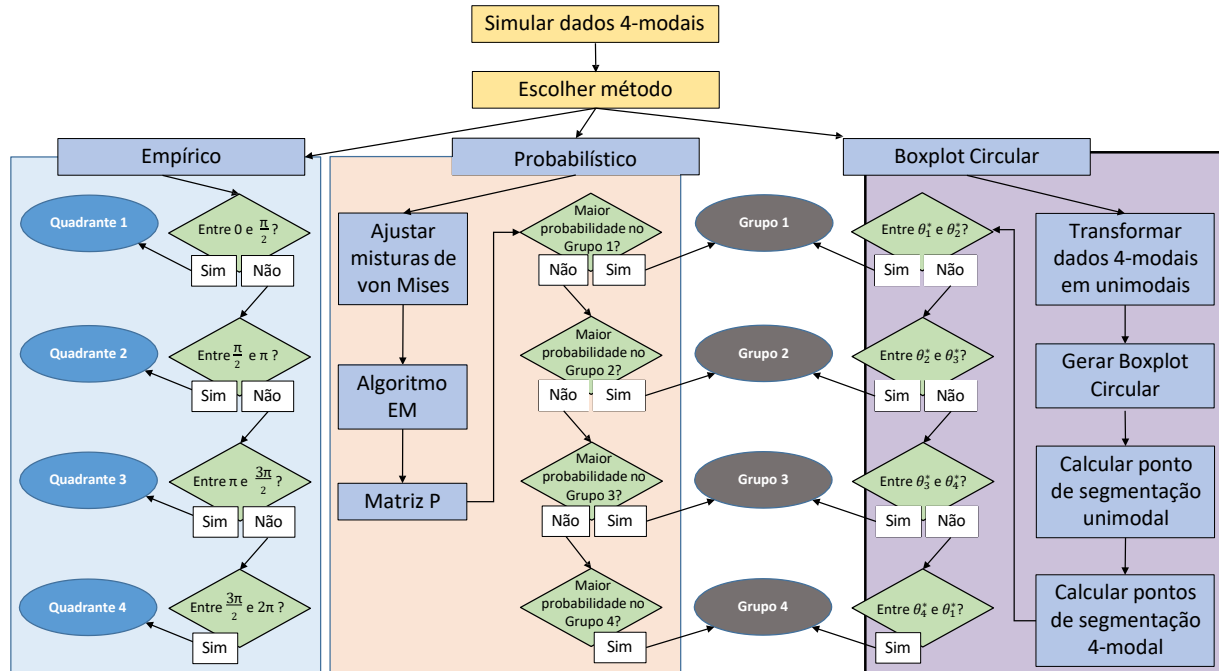
Após essas duas simulações com concentração fixada e médias opostas no círculo, foram simuladas combinações de μ e κ , para estudar o comportamento das decisões do teste de Watson para a distribuição von Mises, em todo o círculo com diferentes concentrações. Para gerar misturas de von Mises, utilizou-se a função `movFM` do pacote `movMF` (Meer; B., 2021).

As misturas de von Mises foram geradas com concentrações κ nos valores 5, 10, 20, 40, 80, 100 e 200, e valores de $\mu = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$, que receberam incrementos de $\pi/32$ até que as médias atingissem, respectivamente, os valores anteriores a $\frac{\pi}{2}, \pi, \frac{3\pi}{2}, 2\pi$, evitando repetições. Para cada combinação de μ e κ , foi gerada uma amostra de tamanho 8500, em que as proporções ω_j , necessárias no modelo de misturas, foram obtidas a partir de $\omega_1 = \frac{5000}{8500}$, $\omega_2 = \frac{2000}{8500}$, $\omega_3 = \frac{1000}{8500}$ e $\omega_4 = \frac{500}{8500}$.

Realizou-se testes de Watson, para adequação à distribuição von Mises em cada um dos quatro quadrantes, ajustados ao círculo completo, totalizando 476 configurações distintas: 17 valores de $\mu \times 7$ valores de $\kappa \times 4$ quadrantes, com 1000 repetições para cada configuração. Para apresentar a quantidade de rejeições, foram construídos gráficos de calor com as decisões do teste de Watson pelo o pacote ggplot2 (Wickham, 2016).

Os métodos de agrupamento para os dados 4-modais utilizados na simulação foram: (i) Método Empírico; (ii) Método Probabilístico por algoritmo EM; e (iii) Método do Boxplot Circular seguindo as etapas apresentadas na Figura 4.1.

Figura 4.1 – Fluxograma das etapas das metodologias para segmentações dos dados 4-modais.



Fonte: Do autor (2025).

5 RESULTADOS E DISCUSSÕES

O estudo de simulação foi composto por duas etapas: simulação de quatro conjuntos de dados com distribuição von Mises de tamanhos n_j diferentes e parâmetros μ e κ ; e várias simulações de misturas de von Mises, em que os parâmetros de ponderação ω_j foram determinados pelas proporções de n_j simuladas na etapa anterior, μ com valores entre 0 e 2π em incrementos de $\frac{\pi}{32}$, e o parâmetro de concentração $\kappa = 5, 10, 20, 40, 80, 100$ e 200.

5.1 Uso do Método Empírico em dados provenientes da distribuição von Mises

Para realizar a verificação da multimodalidade, foram criados quatro conjuntos de dados, supondo distribuição von Mises com parâmetros $\kappa = 4$, considerando tamanhos de amostra n_j , sendo $n_1 = 5000$, $n_2 = 2000$, $n_3 = 1000$ e $n_4 = 500$, com j representado cada quadrante do ciclo trigonométrico. Dois casos serão estudados: $\mu = \pi$ e $\mu = 0$:

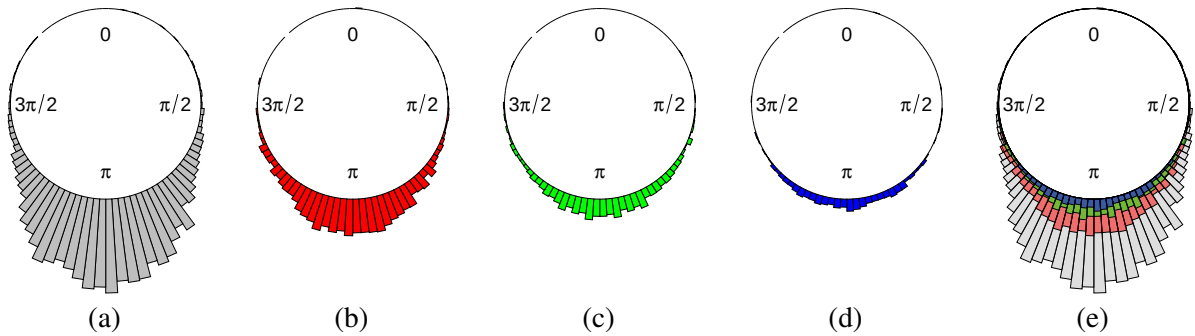
5.1.1 Parâmetro: $\mu = \pi$

Foram gerados quatro conjuntos de dados, supondo distribuição von Mises com parâmetros $\mu = \pi$ e $\kappa = 4$ pela função `rvonmises` do pacote `circular`, no software R (Team, 2025) como mostra a Figura 5.1 considerando tamanhos de amostra n_j , sendo $n_1 = 5000$ (Figura 5.1 (a)), $n_2 = 2000$ (Figura 5.1 (b)), $n_3 = 1000$ (Figura 5.1 (c)) e $n_4 = 500$ (Figura 5.1 (d)) em que cada cor representa um conjunto de dados diferentes. A Figura 5.1 (e) mostra os quatro conjuntos de dados simulados sobrepostos formando um único conjunto de tamanho 8500.

Em seguida, cada conjunto de dados foi ajustado para ocupar um quarto de círculo pela equação (4.1). Essa transformação permite que cada conjunto de dados, apesar de gerados, a partir dos mesmos parâmetros, possam ser distribuídos pelos quatro quadrantes exibindo a multimodalidade, conforme ilustrado na Figura 5.2, gráficos (a) a (d). Na Figura 5.2 (e) apresenta-se a distribuição 4-modal simétrica, adotando-se a formatação em graus com 0° no topo da circunferência e sentido horário por se considerar uma simulação de dados circulares temporais.

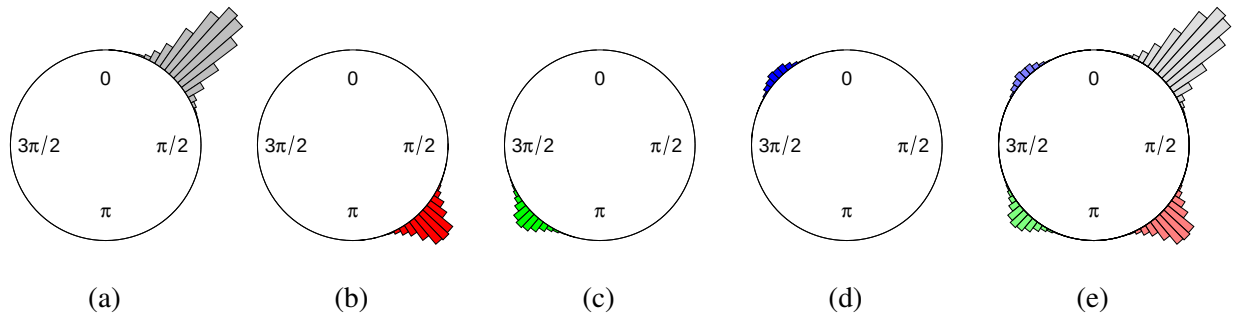
Na Tabela 5.1, apresentam-se as estatísticas correspondentes às Figuras 5.2 (a), (b), (c) e (d), com o tamanho n_j , os pontos mínimos e máximos para cada agrupamento, o arco entre

Figura 5.1 – Histogramas circulares dos dados simulados a partir de von Mises de parâmetros $\mu = \pi$ e $\kappa = 4$: (a) $n_1 = 5000$, (b) $n_2 = 2000$, (c) $n_3 = 1000$, (d) $n_4 = 500$ e (e) $n = 8500$.



Fonte: Do autor (2025).

Figura 5.2 – Histogramas circulares dos dados simulados a partir de $\mu = \pi$ transformados: (a) $n_1 = 5000$, (b) $n_2 = 2000$, (c) $n_3 = 1000$, (d) $n_4 = 500$ e (e) $n = 8500$.



Fonte: Do autor (2025).

os pontos mais próximos dos conjuntos vizinhos e os limites que definem os pontos de segmentação, que no caso empírico são os valores dos arcos quadrantis. Também foram realizados ajustes à distribuição de von Mises para cada um dos quatro conjuntos de dados simulados, e os parâmetros estimados resultantes são exibidos.

Tabela 5.1 – Tamanho (n), valores mínimos (min), máximos (max), arco entre os agrupamentos, limites, média circular (μ), concentração (κ), para os grupos do conjunto de dados simulados com média $\mu = \pi$.

Dados	n	min(°)	max(°)	Arco(°)	Limites(°)	μ (°)	κ
1	5000	2,81	77,76	12,70	90	44,86	54,69
2	2000	90,46	173,35	21,32	180	135,11	53,02
3	1000	194,67	253,30	21,67	270	224,61	53,44
4	500	274,97	351,30	11,51	360	315,01	60,48

Fonte: Do autor (2025).

A análise dos resultados apresentados na Tabela 5.1 revela que cada distribuição von Mises está contida em seu respectivo quadrante, com as 5000 observações simuladas contidas no primeiro quadrante, com a menor delas em $2,81^\circ$ e a maior $77,76^\circ$, e o mesmo para os outros de tamanho 2000, 1000 e 500. Os vales, ou seja, o espaço sem pontos entre os extremos dos agrupamentos são apresentados na coluna *Arco*, indicam uma boa segmentação dos dados, já que os valores extremos dos agrupamentos estão separados por pelo menos $11,51^\circ$, chegando a até $21,67^\circ$ entre os quadrantes 3 e 4.

Os dados foram simulados com distribuição von Mises entre 0 e 2π com um parâmetro de média $\mu = \pi$, que corresponde à média linear de 0 e 2π . Ao serem reduzidos para intervalos de comprimento $\pi/2$, os dados se posicionaram no centro de cada quadrante, como evidenciado pelos valores da coluna μ .

Além disso, os dados foram inicialmente simulados com uma concentração de $\kappa = 4$, mas após a redução para os quadrantes, observou um aumento na concentração, que naturalmente agrupa os dados mais próximos de seus centros angulares, com valores variando em torno de 53,02 e 60,48. A segmentação dos agrupamentos pelo método empírico dividindo em quadrantes coincide com os dados transformados da Figura 5.2.

Para testar se os dados seguem distribuições Uniforme ou von Mises, eles foram multiplicados por quatro para que os valores sejam distribuídos por todo o círculo. Assim as Figuras 5.1 (a), (b), (c) e (d), apresentam os agrupamentos que serão analisados e chamados respectivamente de $G_{\pi;1}^{Emp}$, $G_{\pi;2}^{Emp}$, $G_{\pi;3}^{Emp}$ e $G_{\pi;4}^{Emp}$.

A Tabela 5.1 mostra os resultados do teste de espaçamento de Rao e os testes de Watson para os 4 grupos por quadrantes, para os dados agrupados de forma Unimodal (Figuras 5.1 (e)) conforme recomendado por Landler et al. (2018), e para os dados 4-modais agrupadas no mesmo círculo (Figuras 5.2 (e)).

Nesta simulação, em que as modas estão centradas nos quadrantes, a segmentação intuitiva dos agrupamentos por quadrantes mostra-se adequada, pois os quatro conjuntos gerados rejeitaram a hipótese de uniformidade e não rejeitaram a hipótese de seguir uma distribuição von Mises, fornecendo evidências em favor dessa distribuição. Para o caso 4-modal, a hipótese nula foi rejeitada, já que o teste de Watson avalia a adequação a distribuições unimodais, sendo esperado que ele rejeite uma distribuição 4-modal.

Quadro 5.1 – Decisão dos testes para as amostras simuladas ($\mu = \pi$) considerando os dados separadamente, agrupados sem transformação (Unimodal), agrupados com transformação (4-Modal)

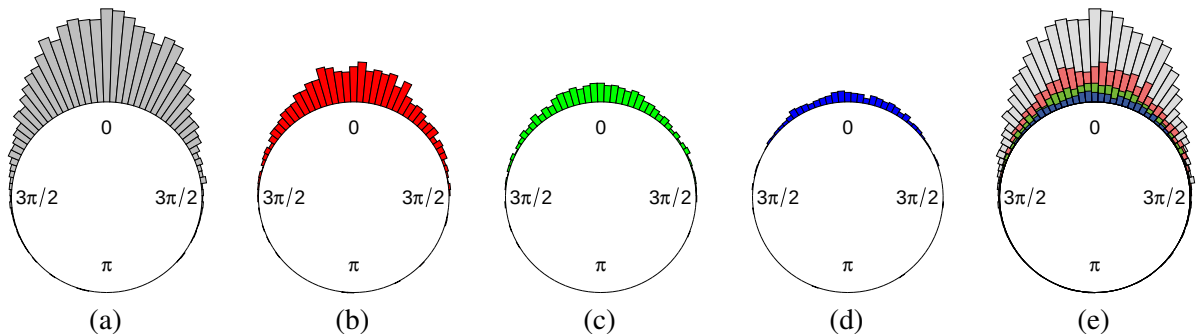
Quadrante (G)	Espaçamento de Rao	Watson	
		Uniforme	Von Mises
$G_{\pi;1}^{Emp}$	Rejeita H_0	Rejeita H_0	Não Rejeita H_0
$G_{\pi;2}^{Emp}$	Rejeita H_0	Rejeita H_0	Não Rejeita H_0
$G_{\pi;3}^{Emp}$	Rejeita H_0	Rejeita H_0	Não Rejeita H_0
$G_{\pi;4}^{Emp}$	Rejeita H_0	Rejeita H_0	Não Rejeita H_0
Unimodal	Rejeita H_0	Rejeita H_0	Não Rejeita H_0
4-Modal	Rejeita H_0	Rejeita H_0	Rejeita H_0

Fonte: Do autor (2025).

5.1.2 Parâmetro: $\mu = 0$

Se os dados simulados anteriormente coincidiram com os quadrantes estatísticos, agora realizou-se uma nova simulação com $\mu = 0$ e $\kappa = 4$, considerando tamanhos de amostra n_j , sendo $n_1 = 5000$ (Figura 5.3 (a)), $n_2 = 2000$ (Figura 5.3 (b)), $n_3 = 1000$ (Figura 5.3 (c)) e $n_4 = 500$ (Figura 5.3 (d)). A Figura 5.3 (e) mostra os quatro conjuntos de dados simulados juntos formando um único conjunto de tamanho 8500, em que cada cor corresponde a um dos conjuntos de dados simulados.

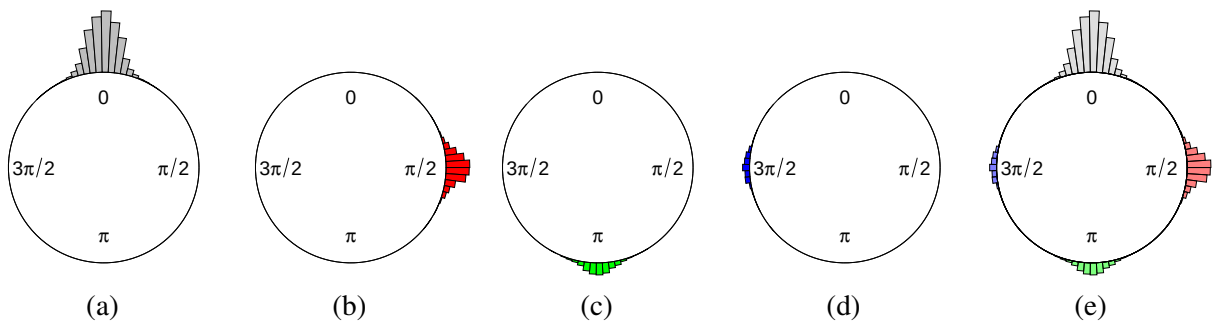
Figura 5.3 – Histogramas circulares dos dados simulados a partir de von Mises de parâmetros $\mu = 0$ e $\kappa = 4$ por quadrantes transformados em ciclos: (a) $n_1 = 5000$, (b) $n_2 = 2000$, (c) $n_3 = 1000$, (d) $n_4 = 500$ e (e) $n = 8500$.



Fonte: Do autor (2025).

Os conjuntos de dados foram reduzidos utilizando a equação (4.1), e estão disponíveis na Figuras 5.4 de (a) a (d). Na Figura 5.4 (f), tem-se uma distribuição 4-modal simétrica, porém agora os agrupamentos não estão mais posicionados em um único quadrante.

Figura 5.4 – Histogramas circulares dos dados simulados a partir de $\mu = 0$ transformados: (a) $n_1 = 5000$, (b) $n_2 = 2000$, (c) $n_3 = 1000$, (d) $n_4 = 500$ e (e) $n = 8500$.



Fonte: Do autor (2025).

A segmentação pelo método empírico está sendo feita por quadrantes, e é possível perceber na Figura 5.4 (e) que cada quadrante tem metades de von Mises. Na Tabela 5.2 é apresentado mais detalhes sobre esses novos quadrantes, revelando discrepâncias entre os tamanhos simulados dos quadrantes e os tamanhos observados após a simulação. Especificamente, os tamanhos dos quadrantes não correspondem aos valores simulados de 5000, 2000, 1000 e 500. Em vez disso, os tamanhos foram 3466, 1527, 737 e 2770, respectivamente, indicando uma diferença alta entre os valores simulados e os agrupados.

Tabela 5.2 – Tamanho (n), valores mínimos (min), máximos (max), arco entre os pontos mais próximos de agrupamentos vizinhos (*Arco*), limites, média circular (μ), concentração (κ), para o conjunto de dados simulados com média $\mu = \pi$.

Dados	n	min(°)	max(°)	Arco(°)	Limites(°)	μ (°)	κ
1	3466	0,01	89,99	0,02	90	25,01	3,46
2	1527	90,01	179,99	0,02	180	120,34	2,90
3	737	180,01	269,84	0,18	270	210,83	2,83
4	2770	270,02	359,99	0,02	360	300,47	7,42

Fonte: Do autor (2025).

Na coluna **Arco**, não há espaço entre os quadrantes com valores que variam entre $0,01^\circ$ e $0,18^\circ$, ou seja, os vales não estão posicionados nos limites quadrantais. Isto sugere que os dados

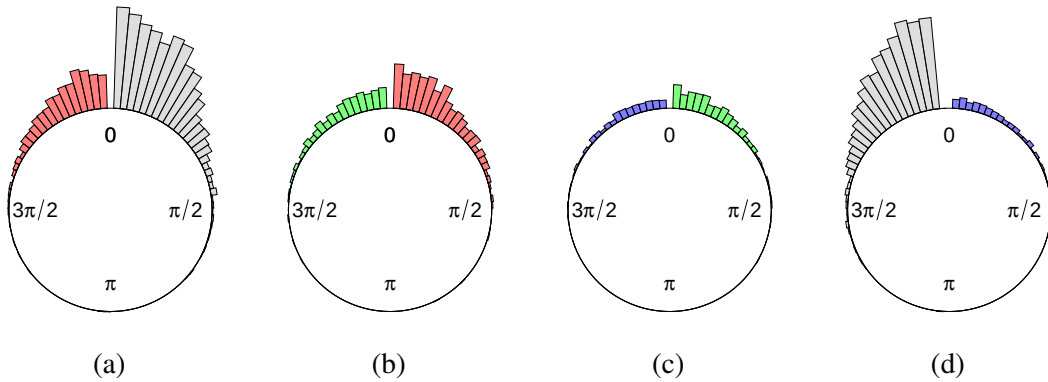
simulados não foram bem agrupados dentro de cada quadrante, resultando em uma sobreposição ou proximidade excessiva entre os limites de cada agrupamento.

Além disso, as médias circulares (μ) observadas para cada quadrante situam-se próximas da metade da von Mises de maior tamanho, sugerindo uma tendência centrada em torno dos pontos de maior concentração de dados.

Os dados foram inicialmente simulados com uma concentração de $\kappa = 4$, mas após a redução para os quadrantes, observou-se uma diminuição na concentração dos três primeiros quadrantes, com valores de κ reduzidos para 3,134, 2,900 e 2,828, respectivamente. Por outro lado, o quarto quadrante, que agrega metades de von Mises com a maior diferença (5000 e 500), apresentou um aumento na concentração, com $\kappa = 7,415$.

Para as análises de adequação as distribuições os dados de cada quadrante foram redistribuídas no círculo completo e nomeadas como $G_{0;1}^{Emp}$, $G_{0;2}^{Emp}$, $G_{0;3}^{Emp}$ e $G_{0;4}^{Emp}$, que são mostrados nas Figuras 5.5 (a), (b), (c) e (d).

Figura 5.5 – Histogramas circulares dos dados simulados a partir de $\mu = 0$ transformados: (a) $n_1 = 5000$, (b) $n_2 = 2000$, (c) $n_3 = 1000$, (d) $n_4 = 500$.



Fonte: Do autor (2025).

Foram realizados os testes de adequação para as distribuições uniforme e von Mises, em que cada linha representa uma das figuras apresentadas na Figura 5.5.

Após o deslocamento, todos os novos ciclos rejeitam a hipótese de seguir uma distribuição von Mises, porém não rejeitam para von Mises ao considerar os ciclos juntos e multiplicados pela quantidade de ciclos, já que mesmo após o deslocamento, os dados continuam sendo gerados por distribuições von Mises de mesma média e concentração, μ e κ , por isso não rejeitaram para o teste de Watson (vM).

Quadro 5.2 – Decisão dos testes para as amostras simuladas ($\mu = 0$) considerando os dados separadamente, agrupados sem transformação (Unimodal), agrupados com transformação (4-Modal)

Quadrante (G)	Espaçamento de Rao	Watson	
		Uniforme	Von Mises
$G_{0;1}^{Emp}$	Rejeita H_0	Rejeita H_0	Rejeita H_0
$G_{0;2}^{Emp}$	Rejeita H_0	Rejeita H_0	Rejeita H_0
$G_{0;3}^{Emp}$	Rejeita H_0	Rejeita H_0	Rejeita H_0
$G_{0;4}^{Emp}$	Rejeita H_0	Rejeita H_0	Rejeita H_0
Unimodal	Rejeita H_0	Rejeita H_0	Não Rejeita H_0
4-Modal	Rejeita H_0	Rejeita H_0	Rejeita H_0

Fonte: Do autor (2025).

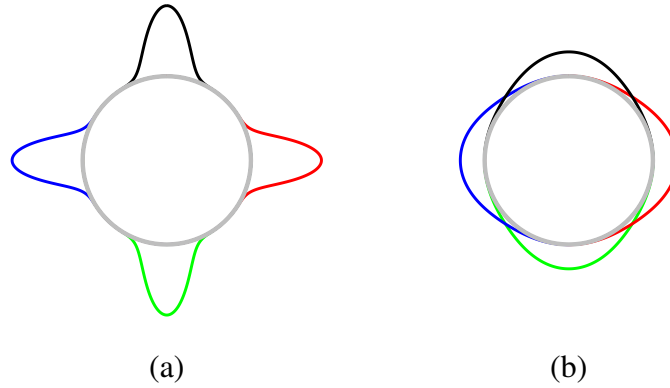
Portanto, faz-se necessário a utilização de um método que forneça uma segmentação adequada para identificar a multimodalidade. Algumas configurações de von Mises foram analisadas e estão apresentadas no item seguinte.

5.2 Estudo da multimodalidade em misturas de distribuição von Mises

Os resultados das simulações apresentadas na subseção 5.1 são provenientes da mesma concentração $\kappa = 4$ no círculo unimodal. No caso em que os agrupamentos ficaram posicionados nos centros dos quadrantes ($\mu = \pi$). As concentrações ficaram entre 53,02 e 60,48 no círculo 4-modal. Esse aumento no valor de κ fez com que as caudas das distribuições de von Mises fiquem contidas em seus quadrantes, ou seja, os agrupamentos não tinham comprimento superior a $\pi/2$ (Figura 5.6 (a)). No entanto se considerarmos valores de κ menores, as curvas individuais terão comprimento maior que $\pi/2$ (Figura 5.6 (b)).

Assim, as simulações anteriores tinham agrupamentos com comprimento menor que $\frac{\pi}{2}$, o que permitiu simular primeiro em um círculo completo e depois reduzir para 1/4 de círculo. Demonstrou-se que combinações de distribuições de von Mises, mesmo compartilhando os mesmos parâmetros, mas em tamanhos diferentes, rejeitam a hipótese nula de adequação à distribuição de von Mises se os agrupamentos tiverem suas segmentações em regiões com alta densidade de dados. Diante disso, para investigar situações como na Figura 5.6 (b), em que

Figura 5.6 – Curvas 4-modais simétricas com concentrações: (a) $\kappa = 40$ e (b) $\kappa = 5$.



Fonte: Do autor (2025).

há interações entre as caudas das curvas, os dados foram gerados, a partir de misturas de von Mises, com diferentes concentrações κ e médias circulares μ , e proporções ω_j conforme 7.1.2.

Dessa forma, para cada κ foram realizadas 64 mil simulações. Para cada simulação foram testados os três métodos de segmentação para verificar os limites de determinação das segmentações

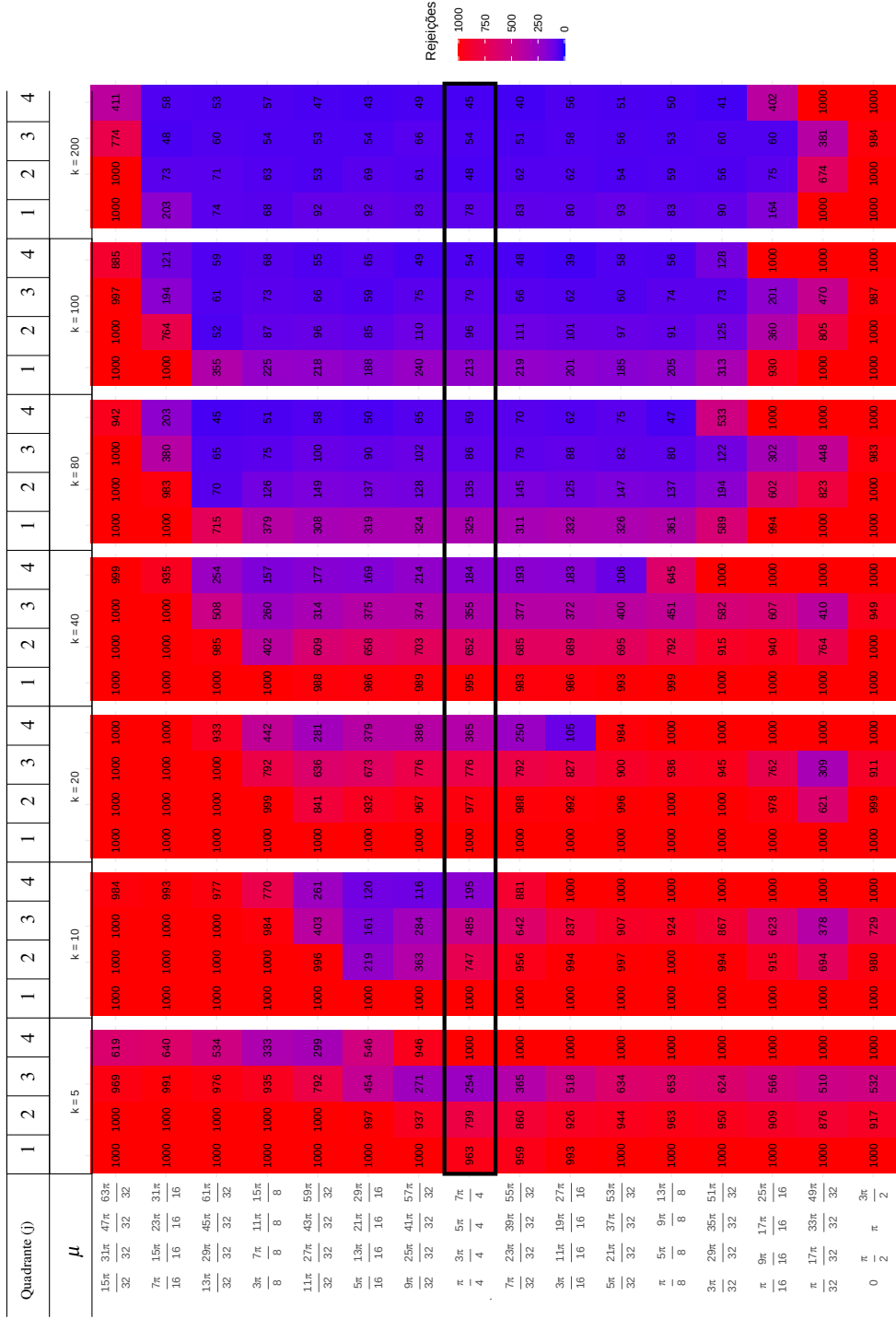
5.2.1 Método Empírico

O primeiro método de segmentação foi o mesmo utilizado na subseção 5.1: o Método Empírico, particionando em quadrantes. Realizou-se testes de Watson para adequação à distribuição de von Mises em cada um dos quatro quadrantes para cada combinação de μ e κ como mostra a Figura 5.3.

Os valores de μ_j se referem as posições das modas de tamanho n_j , por exemplo, o valor de μ_3 na última linha é de π o que significa que a média circular da von Mises de tamanho 1000 (n_3) é $\mu_3 = \pi$. Assim, em 1000 testes realizados para $\kappa = 5$ ocorreram 532 rejeições à hipótese nula do teste de Watson para a distribuição von Mises para os dados posicionados no terceiro quadrante quando $\mu_3 = \pi$. Seja $G(\mu, \kappa, j)$ o grupo que está sendo testado, o valor em cada célula do mapa de calor corresponde ao número de rejeições de $G(\mu, \kappa, j)$ assim, por exemplo, o número de rejeições em $G(\pi, 5, 3) = 532$.

O mapa de calor evidencia que o valor de κ é determinante na quantidade de rejeições pelo círculo, quando κ é pequeno, como nas colunas de $\kappa = 5$ até 20 e o mapa é predominantemente vermelho, ou seja, tem um elevado número de rejeições da hipótese nula (H_0). Nessas concentrações mais baixas como apresentado na Figura 5.6 (b) há interações entre as curvas, e

Tabela 5.3 – Mapa de calor por rejeição do teste de Watson às distribuições von Mises agrupados por quadrante com parâmetro $\kappa = 5, 10, 20, 40, 80, 100$ e 200 , e μ de 0 a 2π em incrementos de $\frac{\pi}{32}$.



Fonte: Do autor (2025).

portanto independente das modas estarem posicionadas nos centros dos quadrantes, ou não, o teste de Watson rejeita a hipótese nula pois as caudas das curvas vizinhas interferem no reconhecimento da distribuição von Mises nos dados do quadrante em análise.

A partir de $\kappa = 40$ é possível observar uma diminuição nas rejeições das regiões centrais do mapa de calor, que são aquelas em que as médias circulares estão posicionadas no centro do quadrante, próximos de $\mu_1 = \frac{\pi}{4}$ e seus equivalentes nos outros quadrantes, indicando que nessa concentração as interferências de dados gerados a partir de von Mises de μ vizinho começam a perder o efeito.

Para $\kappa = 80, 100$ e 200 , o mapa de calor se torna predominantemente azul, e as exceções passam a ser as regiões em que as médias circulares simuladas estão nos arcos quadrantis ou próximas a eles. Isso ocasiona uma segmentação irregular dos dados entre os grupos, tal como apresentado na Figura 5.5, gerando um número maior de rejeições como esperado.

Destaca-se também que há um "movimento" da diminuição de rejeição da esquerda para a direita, nas sub-colunas dos quadrantes j . Esse fato se deve ao número diferenciado da quantidade de dados gerados n_j , que justifica o fato de o vermelho ser mais intenso quando $j = 1$, pois em $n_1 = 5000$ os vizinhos são inferiores, $n_2 = 2000$ e $n_4 = 500$. Esta modificação do tamanho amostral permite também verificar a ocorrência de multimodalidade não equiprovável em ω .

Desse modo, o Método Empírico só terá baixos valores de rejeição se toda a região modal estiver contida em um único quadrante. Para essa simulação, essas regiões serão as próximas das médias circulares: $\mu_1 = \frac{\pi}{4}$, $\mu_2 = \frac{3\pi}{4}$, $\mu_3 = \frac{5\pi}{4}$ e $\mu_4 = \frac{7\pi}{4}$, que estão destacadas na Figura 5.3. Quanto maior for o κ , maior será a probabilidade de que toda a região modal fique compreendida dentro de um único quadrante.

5.2.2 Método Probabilístico por algoritmo EM

Ao realizar simulações agrupando os dados 4-modais pelo Método Probabilístico utilizando o algoritmo EM, observa-se pela Figura 5.4 que as médias circulares μ_j não determinam a quantidade de rejeições. Isso é evidenciado pelo fato de que dentro de cada coluna não há variação de coloração. Tal observação sugere que, neste método, não é necessário que o ângulo coincida exatamente com o centro de um agrupamento "empírico".

Esse fato pode ser comparado quando verificam se as médias circulares $\mu_1 = \frac{\pi}{4}$, $\mu_2 = \frac{3\pi}{4}$, $\mu_3 = \frac{5\pi}{4}$ e $\mu_4 = \frac{7\pi}{4}$ destacadas no retângulo preto das Figuras 5.4 e 5.3. Nas duas linhas destaca-

das os valores de $G(\mu, \kappa, j)$, em sua maioria, são bem parecidos. Para o Método Empírico essa linha destacada tem as menores quantidades de rejeições, para o Método Probabilístico esses bons resultados ocorrem para todas as outras médias circulares.

Isso acontece pelo fato do Método Probabilístico agrupar os dados, a partir dos valores estimados dos parâmetros de von Mises encontrados nas modelagem da mistura. Esse método tem a sensibilidade de agrupar cada observação θ pela sua proximidade com a média circular estimada, e considerando o quão dispersos essas observações devem estar a partir do valor estimado de concentração.

As concentrações κ desempenham um papel fundamental. No método empírico, foi observado que baixas concentrações resultavam em um maior número de rejeições da hipótese nula; no método atual, essa premissa continua válida, mas agora, sem o efeito das médias circulares μ_j , é possível afirmar que altas concentrações levam a um menor número de rejeições da hipótese nula. Assim, pode-se concluir que o Método Probabilístico por algoritmo EM é eficaz, para determinar agrupamentos em dados 4-modais, independentemente da localização das médias circulares, especialmente em situações de alta concentração.

5.2.3 Método Boxplot Circular

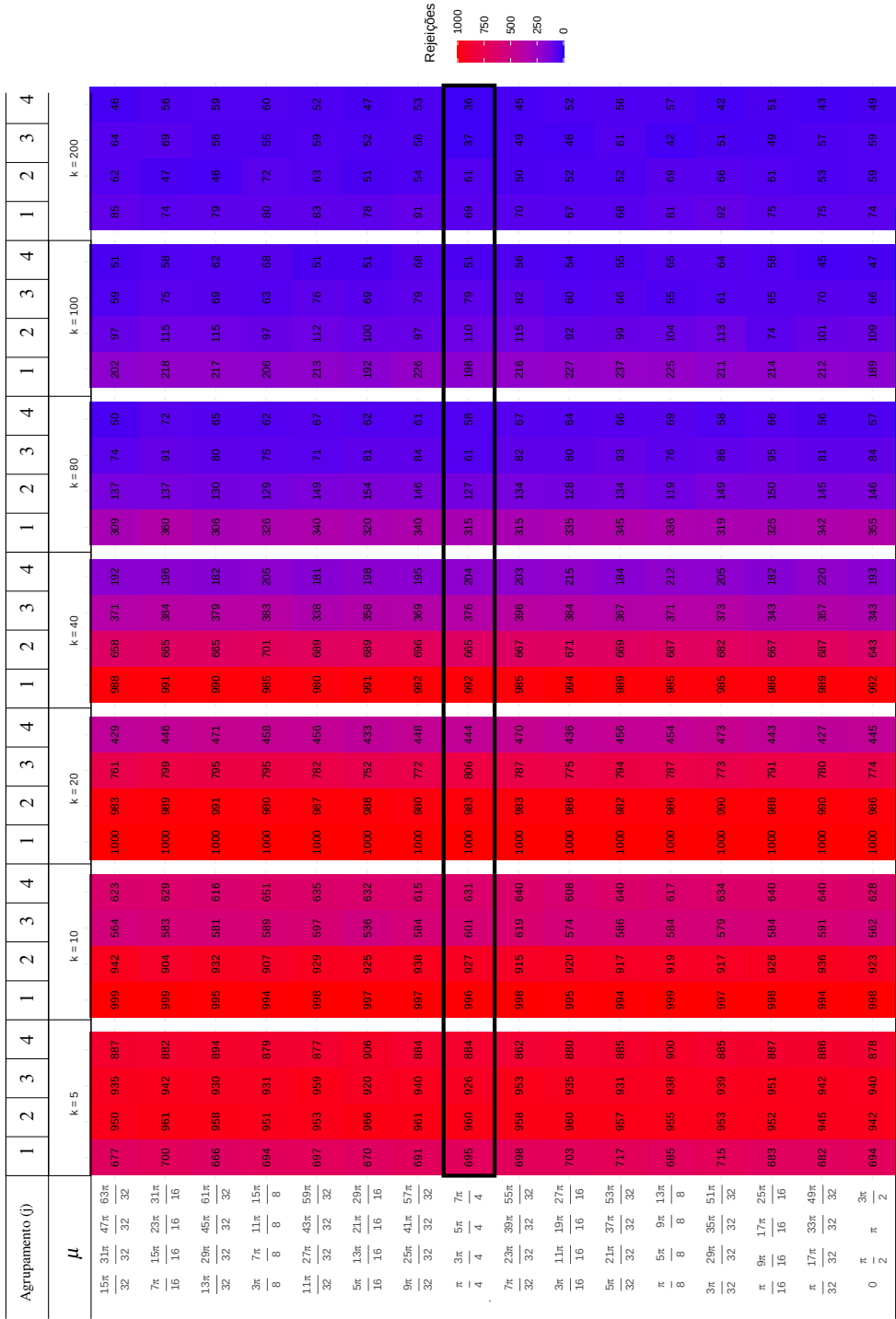
O Método de Segmentação por Boxplot Circular foi aplicado para os dados simulados 4-modais, realizou-se testes de Watson para adequação à distribuição de von Mises em cada um dos quatro quadrantes para cada combinação de μ e κ , conforme ilustrado na Figura 5.5.

O mapa de calor da Figura 5.5 é bem parecido com as rejeições obtidas pelo Método de Segmentação Probabilístico, com κ sendo mais uma vez um fator crucial para a quantidade de rejeições. Para valores baixos de κ , como 5 e 10, observa-se uma predominância de rejeições (cor vermelha), sugerindo uma maior dificuldade em ajustar os dados à distribuição von Mises pela sobreposição entre as distribuições dos quadrantes.

Com o aumento de κ para 40 e acima, há uma redução notável nas rejeições, especialmente nos quais as médias circulares estão próximas ao centro dos quadrantes, indicando que a concentração alta facilita a segmentação dos grupos e a conformidade com a distribuição von Mises.

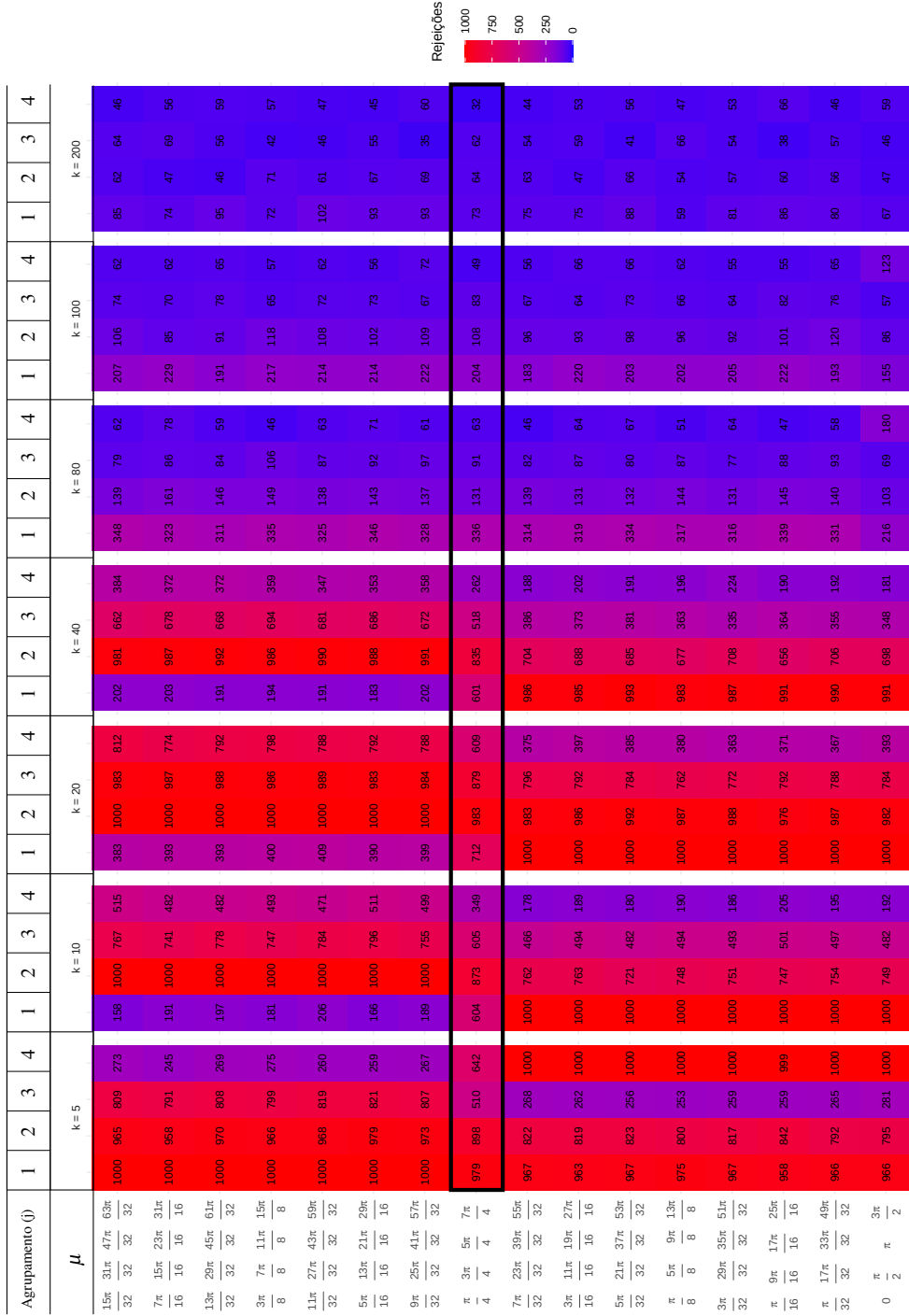
Destaca-se neste método, que, diferentemente do ocorrido no probabilístico (Figura 5.4) há uma modificação do número de rejeição dentro das colunas, que se deve justamente ao

Tabela 5.4 – Mapa de calor por rejeição do teste de Watson às distribuições von Mises agrupados pelo Método Probabilístico com parâmetro $\kappa = 5, 10, 20, 40, 100$ e 200 , e μ de 0 a 2π em incrementos de $\frac{\pi}{32}$.



Fonte: Do autor (2025).

Tabela 5.5 – Mapa de calor por rejeição do teste de Watson às distribuições von Mises, segmentado pelo Método Boxplot Circular com $\kappa = 5, 10, 20, 40, 100, 200$ e μ de 0 a 2π em incrementos de $\frac{\pi}{32}$.



Fonte: Do autor (2025).

deslocamento proposto pelo método. No entanto, percebe-se que de maneira geral, a eficiência é semelhante a ocorrida no Método Probabilístico.

Este comportamento reforça que o Método de Segmentação por Boxplot Circular é eficiente para particionar dados 4-modais, especialmente quando se trata de dados com alta concentração, nos quais a separação entre quadrantes é mais clara.

Diante do exposto, é mostrado na Tabela 5.6 um comparativo do percentil de rejeições entre o Método Empírico, Método Probabilístico por algoritmo EM e Método de Segmentação por Boxplot Circular de acordo com os valores de concentração estudados, $\kappa = 5, 10, 20, 40, 80, 100$ e 200 e a porcentagem de rejeição total, considerando todos os testes realizados nas Figuras 5.3, 5.4 e 5.5.

Tabela 5.6 – Porcentagens das rejeições de H_0 para diferentes valores de κ , nos três métodos estudados, com a soma das linhas.

Método	5	10	20	40	80	100	200	Total
Empírico	84,30%	83,40%	86,64%	70,25%	38,34%	32,43%	21,93%	59,61%
Prob.	86,75%	78,31%	80,44%	55,74%	15,31%	11,02%	5,98%	47,65%
Boxplot	76,02%	60,87%	79,03%	55,75%	15,33%	11,02%	6,09%	43,44%

Fonte: Do autor (2025).

Nota-se que o Método Empírico é o que tem a menor eficiência em todos os casos, exceto para $\kappa = 5$ em que o método com maior percentual de rejeição foi o probabilístico. Para $\kappa = 10$ já foi observada uma melhora no desempenho do Método Probabilístico, mas ainda tem eficiência inferior ao Método Boxplot Circular. A partir de $\kappa = 20$ percebe-se uma variação desprezível entre os percentis obtidos nos dois últimos métodos. Assim, pode-se dizer que o método proposto nesse trabalho, é tão eficiente quanto o Método Probabilístico para altos valores de κ e melhor para valores menores de κ apesar das rejeições serem superiores a 60%.

Na coluna dos totais a maior porcentagem, com alguma distância, foi a do Método Empírico com 59,61%, e as outras duas ficaram próximas, com o Método Boxplot Circular sendo o menor com 43,44% seguido do Método Probabilístico com 47,65%.

Ressalta-se ainda que o Método de Segmentação por Boxplot Circular é também muito mais simples no entendimento e execução já demanda apenas da determinação dos valores dos quartis circulares e não de algoritmos mais complexos como o EM no caso probabilístico. Pode-se dizer que, de certa forma, o método proposto nesse trabalho é um intermediário, em relação

a dificuldade pois tem um fator visual de construção que pode facilitar a interpretação, mas não é tão instintivo quanto o empírico pois traz em sua metodologia o uso de estatísticas descritivas que facilitam a utilização em relação ao Método Probabilístico.

6 CONCLUSÕES

Concluiu com a simulação que o Método Empírico mostra-se menos eficiente na maioria dos casos, exceto para $\kappa = 5$, no qual o Método Probabilístico teve maior rejeição. Com $\kappa = 10$, o Método Probabilístico melhora, mas ainda é menos eficiente que o Método Boxplot Circular. Para $\kappa = 20$, a diferença entre os métodos probabilístico e o proposto diminui, indicando eficiência semelhante para altos valores de κ , mas melhor para valores menores. O Método de Segmentação por Boxplot Circular é simples, exigindo apenas a determinação de quartis circulares, contrastando com a complexidade do algoritmo EM do método probabilístico. O método proposto oferece uma visualização facilitadora da interpretação, situando-se entre a simplicidade do empírico e a complexidade do probabilístico, utilizando estatísticas descritivas para facilitar a aplicação.

REFERÊNCIAS

- ANDERSON, C. M. **Graphical methods for circular and cylindrical data**. [S.l.], 1993.
- BANERJEE, A. *et al.* Clustering on the unit hypersphere using von mises-fisher distributions. **Journal of Machine Learning Research**, v. 6, p. 1345–1382, 2005.
- BOUYEYRON, C.; BRUNET-SAUMARD, C. Model-based clustering of high-dimensional data: A review. **Computational Statistics & Data Analysis**, Elsevier, v. 71, p. 52–78, 2014.
- BRUZZONE, L.; PRIETO, D. F. An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images. **IEEE Transactions on image processing**, IEEE, v. 11, n. 4, p. 452–466, 2002.
- BUTTARAZZI, D. **Package ‘bpDir’**. [S.l.], 2022. R package version 1.0.0.
- BUTTARAZZI, D.; PANDOLFO, G.; PORZIO, G. C. **A boxplot for circular data**. [S.l.]: Oxford, 2018.
- CHAKRABORTY, S.; HAZARIKA, P. J.; ALI, M. M. A multimodal skewed extension of normal distribution: its properties and applications. **Statistics**, Taylor & Francis, v. 49, n. 4, p. 859–877, 2014.
- CHANG-CHIEN, S.-J.; HUNG, W.-L.; YANG, M.-S. On mean shift-based clustering for circular data. **Soft Computing**, Springer, v. 16, p. 1043–1060, 2012.
- CHASANI, P.; LIKAS, A. The uu-test for statistical modeling of unimodal data. **Pattern Recognition**, Elsevier, v. 122, p. 108272, 2022.
- CHASANI, P.; LIKAS, A. Statistical modeling of univariate multimodal data. **arXiv preprint arXiv:2412.15894**, 2024.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. **Journal of the royal statistical society: series B (methodological)**, Wiley Online Library, v. 39, n. 1, p. 1–22, 1977.
- DONNER, A.; GOLDENSHLUGER, A. Statistical inference in circular structural model and fitting circles to noisy data. **Journal of Computational and Graphical Statistics**, Taylor & Francis, p. 1–20, 2024.
- HORNIK, K.; LEISCH, F. **movMF: Mixture of von Mises-Fisher Distributions**. [S.l.], 2014. R package version 1.2.0.
- KATYAR, A.; PANDEY, C. A review of solar radiation models—part i. **Journal of Renewable Energy**, Wiley Online Library, v. 2013, n. 1, p. 168048, 2013.
- KESEMEN, O.; TEZEL, ; ÖZKUL, E. Fuzzy c-means clustering algorithm for directional data (fcm4dd). **Expert Systems with Applications**, Elsevier, v. 58, p. 76–82, 2016.
- LANDLER, L.; RUXTON, G. D.; MALKEMPER, E. P. Circular data in biology: advice for effectively implementing statistical procedures. **Behavioral ecology and sociobiology**, Springer, v. 72, p. 1–10, 2018.

LUND, U.; AGOSTINELLI, C.; AGOSTINELLI, M. C. **Package ‘circular’**. [S.l.], 2017. v. 775, n. 5.

MARDIA, K. V.; JUPP, P. E. **Directional Statistics**. [S.l.]: Wiley, 2000.

MARKOVIĆ, I.; PETROVIĆ, I. Bearing-only tracking with a mixture of von mises distributions. In: IEEE. **2012 IEEE/RSJ International Conference on Intelligent Robots and Systems**. [S.l.], 2012. p. 707–712.

MEER, L. J. P. van der; B., K. M. **movMF: Mixture of von Mises-Fisher Distributions**. [S.l.], 2021. R package version 1.0.0.

MISES, R. von. Über die “ganzzahligkeit” der atomgewichte und verwandte fragen. In: SHEN, A. (Ed.). **Selected Papers of Richard von Mises**. [S.l.]: Dover Publications, 1918. p. 58–80.

QARMALAH, N. M.; EINBECK, J.; COOLEN, F. P. K-boxplots: An extension of boxplot for circular data. **Statistical Papers**, Springer, v. 57, n. 4, p. 975–990, 2016.

RODRÍGUEZ, C. E.; NÚÑEZ-ANTONIO, G.; ESCARELA, G. A bayesian mixture model for clustering circular data. **Computational Statistics & Data Analysis**, Elsevier, v. 143, p. 106842, 2020.

TEAM, C. R. **R: A Language and Environment for Statistical Computing**. [S.l.], 2025. R Foundation for Statistical Computing, Vienna, Austria.

WANG, F. **Space and Space-Time Modeling of Directional Data**. Tese (PhD Dissertation) — Duke University, 2013.

WICKHAM, H. **ggplot2: Create Elegant Data Visualisations Using Grammar of Graphics**. [S.l.], 2016. R package version 3.4.0.

Artigo 2 - Elaborado conforme a norma NBR 6022	
Título do artigo:	Segmentação dos dias da semana em dados de acidentes de trânsito nas rodovias mineiras
Autores:	Luiz Fernando Silva Resende Carla Regina Guimarães Brighenti Clodoaldo Teodosio Santana da Silva



**ARTIGO 2 – SEGMENTAÇÃO DOS DIAS DA SEMANA EM DADOS DE
ACIDENTES DE TRÂNSITO NAS RODOVIAS MINEIRAS
SEGMENTATION OF DAYS OF THE WEEK IN TRAFFIC ACCIDENT DATA ON
MINAS GERAIS HIGHWAYS¹**

Luiz Fernando Silva Resende

<https://orcid.org/0000-0003-4697-6503>

Carla Regina Guimarães Brighenti

<https://orcid.org/0000-0002-7822-3744>

Clodoaldo Teodosio Santana da Silva

<https://orcid.org/0000-0003-4086-1895>

Resumo: Analisou-se a distribuição temporal de acidentes de trânsito nas rodovias de Minas Gerais, com dados fornecidos pela Polícia Militar de Minas Gerais, referentes aos anos de 2020 a 2022, nas regiões de Lavras, São João del Rei e Teófilo Otoni. O objetivo é identificar períodos de maior incidência de acidentes e segmentar os dados por dias da semana, utilizando métodos estatísticos para dados circulares: Empírico, Algoritmo EM e Boxplot Circular. A análise revelou alta dispersão nos horários, com picos de ocorrência entre 6h e 24h, e maior concentração em determinados dias, como segunda-feira e sábado. O Método Boxplot Circular foi o método mais adequado, equilibrando simplicidade e precisão, oferecendo subsídios para políticas públicas de segurança viária.

Palavras-chave: estatística circular; multimodalidade; distribuição von Mises; Algoritmo EM; teste de Watson.

Abstract: The temporal distribution of traffic accidents on Minas Gerais highways was analyzed using data provided by the Minas Gerais Military Police from 2020 to 2022, covering the regions of Lavras, São João del Rei, and Teófilo Otoni. The objective was to identify periods of higher accident incidence and segment the data by days of the week, employing circular statistical methods: Empirical, EM Algorithm, and Circular Boxplot. The analysis revealed high temporal dispersion, with peaks between 6 AM and midnight, and greater concentration on days like Monday and Saturday. The Circular Boxplot method was the most suitable, balancing simplicity and precision, providing support for public road safety policies.

Keywords: circular statistics; multimodality; von Mises distribution; EM Algorithm; Watson test.

¹ Artigo científico apresentado à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária.

1 INTRODUÇÃO

Acidentes automobilísticos têm sido causa de preocupação, pois além das perdas de vidas, causam problemas sociais e econômicos, que vão desde afastamentos do trabalho, invalidez e outros relacionados com a vítima e seus parentes (Marín; Queiroz, 2000).

Em acidentes que envolvem vítimas, quanto mais rápido for o atendimento menores poderão ser os riscos, pois a demora do atendimento pode levar a consequências como o agravamento das condições de saúde dos envolvidos no acidente e até a morte. Um fator que pode agilizar o atendimento é investigar o horário em que o acidente ocorre, já que eles podem ser mais comuns em determinados horários e períodos do dia, dependendo da região (Ladeira; Barreto, 2008).

Os horários de ocorrência de acidentes podem variar de acordo a região. Um estudo numa província no Irã, verificou que nos horários de 12h e 16h havia uma incidência maior de acidentes fatais (Seyed *et al.*, 2013). No Brasil, no estado de São Paulo Deslandes (1999) verificou, ao analisar a distribuição dos atendimentos por hora de entrada em hospitais, que nos horários das 8h às 20 h havia uma demanda maior, exceto nos finais de semana, quando o cenário se inverte, o período com maior número de atendimentos é das 20 às 8 horas. No sudeste da Ásia, verificou-se que, a maior parte dos acidentes ocorriam entre os horários das 18h e 0h (Khare *et al.*, 2012).

A análise dos horários de ocorrência de acidentes de trânsito é fundamental para o desenvolvimento de estratégias de prevenção e alocação eficiente de recursos emergenciais. Este estudo tem como propósito investigar a distribuição temporal desses eventos ao longo do dia e da semana em dados de horários de acidentes de trânsito fornecidos pela Polícia Militar de Minas Gerais (PMMG), utilizando métodos estatísticos para dados circulares que superem as limitações de abordagens tradicionais, as quais frequentemente não consideram a natureza cíclica dos dados horários.

A pesquisa busca identificar períodos de maior incidência de acidentes e segmentar os dados empregando os métodos Empírico, Probabilístico por Algoritmo EM e Boxplot Circular, com o objetivo de avaliar sua eficácia na modelagem da distribuição von Mises e fornecer subsídios para políticas públicas de segurança viária.

2 REFERENCIAL TEÓRICO

A análise de multimodalidade em dados circulares é essencial para compreender a estrutura de conjuntos de dados complexos. Identificar se os dados provêm de uma distribuição com um único pico ou múltiplos picos é crucial para modelagem estatística (Chasani; Likas, 2022). Em dados lineares, a detecção de modas e vales (regiões com baixa densidade de pontos) auxilia na identificação de agrupamentos (Chasani; Likas, 2024). Para dados circulares, a abordagem envolve, após detectar multimodalidade, transformar os dados em uma distribuição unimodal para facilitar análises estatísticas.

O processo inicia-se com a definição dos limites de segmentação para cada grupo que contém uma moda. Três métodos são apresentados neste trabalho: Empírico, Probabilístico (via algoritmo EM) e Boxplot Circular. Após a segmentação, os dados de cada grupo são testados quanto à adequação à distribuição von Mises utilizando o teste de Watson. A seguir, descrevem-se os métodos.

2.1 Método Empírico

O Método Empírico define os pontos de segmentação de forma arbitrária, sem considerar as características intrínsecas dos dados. Por exemplo, em dados temporais, os limites podem ser baseados em convenções, como o início de um dia às 00h00min e o fim às 23h59min. O algoritmo é descrito como:

- a) determinar o número de modas M (grupos).
- b) calcular os M pontos de segmentação θ_{Emp}^* :

$$\theta_{Emp}^* = \theta_I^* + (m - 1) \frac{360^\circ}{M}, \quad \text{para } m = 1, 2, \dots, M, \quad (2.1)$$

em que θ_I^* é o ponto inicial, geralmente adotado como 0° .

Por exemplo, para dados influenciados pelos dias da semana ($M = 7$), os intervalos seriam:

- a) domingo: $[0.00^\circ; 51.43^\circ[$;
- b) segunda: $[51.43^\circ; 102.86^\circ[$;
- c) terça: $[102.86^\circ; 154.29^\circ[$;

- d) quarta: $[154.29^\circ, 205.71^\circ[$;
- e) quinta: $[205.71^\circ, 257.14^\circ[$;
- f) sexta: $[257.14^\circ, 308.57^\circ[$;
- g) sábado: $[308.57^\circ, 360.00^\circ[$.

Ou seja, cada intervalo tem duração de $\frac{360^\circ}{7}$.

2.2 Método Probabilístico por Algoritmo EM

O algoritmo EM, proposto por Dempster et al. (1977) e implementado no pacote `movMF` do R (Meer; B., 2021), é utilizado para segmentação probabilística. Ele alterna entre as etapas de Expectativa (E) e Maximização (M) até a convergência (Bruzzzone; Prieto, 2002). O algoritmo modela os dados como uma mistura de distribuições von Mises, utilizando as probabilidades P_{nk} de cada ponto x_n pertencer a um agrupamento $k = 1, \dots, K$. As etapas são:

- a) expectativa (E): Calcular a probabilidade P_{nk} de cada ponto x_n pertencer ao agrupamento k :

$$P_{nk} = \frac{\omega_k f(x_n | \mu_k, \kappa_k)}{\sum_{j=1}^K \omega_j f(x_n | \mu_j, \kappa_j)}, \quad (2.2)$$

em que ω_k é a proporção do grupo k , e f é a densidade da distribuição von Mises vM avaliada em x_n .

- b) maximização (M): Atualizar os parâmetros μ_k , κ_k e ω_k para maximizar a função de verossimilhança $\mathcal{L}(\mu, \kappa, \pi)$:

$$\mathcal{L}(\mu, \kappa, \omega) = \sum_{n=1}^N \sum_{k=1}^K P_{nk} \log [\omega_k f(x_n | \mu_k, \kappa_k)]. \quad (2.3)$$

As probabilidades P_{nk} são calculadas com o pacote `movMF` (Hornik; Leisch, 2014), que ajusta modelos de mistura de von Mises. Cada ponto é atribuído ao agrupamento com maior probabilidade (Chasani; Likas, 2024).

2.3 Método Boxplot Circular

O Método Boxplot Circular, utiliza quartis circulares (Buttarazzi; Pandolfo; Porzio, 2018) para segmentar dados multimodais simétricos, comuns em dados temporais (e.g., dias

da semana). Ele é mais simples que o método probabilístico, pois não exige ajustes iterativos complexos, aproveitando visualizações gráficas intuitivas. O algoritmo é:

- a) determinar o número M de grupos.
- b) transformação de dados multimodais para unimodais para dados com M -modas simétricos, conforme Landler et al. (2018):
 - multiplicar os ângulos dos dados por M .
- c) construção do boxplot circular, determinando:
 - L_A (limite anti-horário): tipicamente, $Q_A - 1.5 \times IQR$;
 - L_H (limite horário): tipicamente, $Q_H + 1.5 \times IQR$, na qual $IQR = Q_H - Q_A$.
- d) cálculo do ponto de segmentação $\theta_{L_i}^*$ para dados unimodais.

Os dados foram testados, para uniformidade e distribuição von Mises, utilizando o teste de espaçamento de Rao, teste de Watson, para a distribuição uniforme e teste de Watson para a distribuição von Mises, respectivamente, pelas funções `rao.test` e `watson.test` do pacote `circular` (Lund; Agostinelli; Agostinelli, 2017), ao nível de significância de 5%. A análise da primeira simulação foi repetida para a segunda, e os resultados foram comparados.

3 METODOLOGIA

Analizou-se um conjunto de dados de horários de acidentes de trânsito fornecidos pela Polícia Militar de Minas Gerais (PMMG). O banco de dados inclui registros diários de acidentes ocorridos nas regiões de Lavras, São João del Rei e Teófilo Otoni durante os anos de 2020 a 2022. Os horários dos acidentes foram convertidos em ângulos θ_i , representando as 24 horas do dia, distribuídas ao longo de um círculo trigonométrico, de forma que 1 hora corresponde a 15° , e um dia completo é representado por 360° .

Identificou-se a presença de arredondamentos nos registros dos horários dos acidentes, concentrados em horários de hora exata e meia hora (30 minutos). Os dados passaram por um processo de filtragem para eliminar valores incoerentes, garantindo maior precisão na análise. O estudo prosseguiu com os dados filtrados que foram inicialmente agrupados por horários e posteriormente organizados de acordo com os dias da semana em que os acidentes ocorreram, evidenciando cada dia da semana como um agrupamentos no diagrama de rosa.

O padrão de horários considerando inicialmente referente aos sete dias da semana considerados inicialmente, iniciando às 00h e encerrando às 23h59m, visou determinar a presença de multimodalidade e os limites entre os três métodos: (i) Método Empírico, (ii) Método Probabilístico por algoritmo EM e (iii) Método por Boxplot Circular.

Para a análise dos dados, utilizou-se o software R (Team, 2025) os pacotes `bpDir` (Buttarazzi, 2022) para construção do boxplot circular, `circular` (Lund; Agostinelli; Agostinelli, 2017) para o cálculo das estatísticas circulares, testes de uniformidade, adequação a von Mises e homogeneidade entre duas amostras. Para estimar os parâmetros das multiplas von Mises no círculo e a matriz de probabilidade P_{nk} utilizou-se o `movMF` (Hornik; Leisch, 2014).

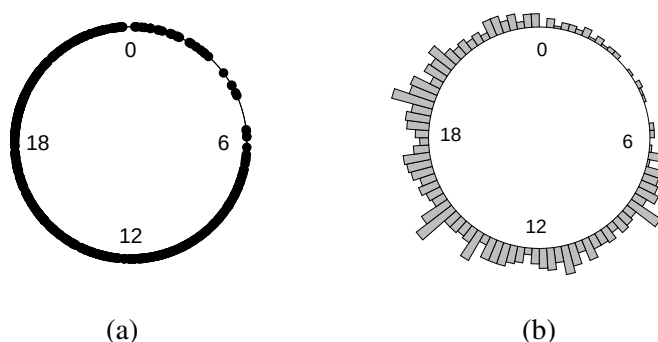
Para determinar a quantidade dos agrupamentos nos dados com os horários de acidentes de trânsito, inicialmente no método empírico foi estabelecido um agrupamento para cada dia da semana por meio de análises gráficas, totalizando sete grupos. A construção dos gráficos de dispersão circular e histogramas circulares (diagramas de rosas) para ilustrar a distribuição dos dados foi realizada com o pacote `ggplot2` (Wickham, 2016).

Para o ajuste realizado nas misturas de von Mises, a quantidade de grupos foi determinada pelos modelos que apresentaram menores valores de AIC e BIC (Mooney; Helms; Jolliffe, 2003).

4 RESULTADOS E DISCUSSÕES

Utilizou-se, na análise a seguir, o registro de 601 horários de acidentes nas rodovias horários desconsiderando os dias da semana do registro da ocorrência. Nas Figuras 4.1 (a) e (b), são exibidas, respectivamente, o diagrama de dispersão circular de pontos e o histograma circular.

Figura 4.1 – Representações circulares dos horários de ocorrência dos acidentes. (a) Diagrama de dispersão circular; (b) Histograma circular.



Fonte: Do autor (2025).

Na Figura 4.1 (a) percebe-se que o único período com ausência de acidentes é durante a madrugada, entre meia-noite e seis horas, diferente do restante do dia, em que todos os horários apresentam ocorrência de acidentes. Na Figura 4.1 (b) observa-se uma representação mais clara das frequências de acidentes entre as 6 horas e meia noite que não era possível de se perceber no diagrama de dispersão circular, com o maior pico sendo no início da manhã por volta das 8 horas, e no decorrer do dia percebe-se algumas oscilações, com fluxos de acidentes que aumentam e diminuem, com destaque para o aumento desse fluxo após as 18 horas.

Obteve-se uma média circular dos horários de ocorrência de acidentes igual a 231° ou 15h24min. No entanto, ao contrário do que poderia ser sugerido apenas pela média circular, que indica uma concentração de acidentes no período da tarde, o histograma revela que os acidentes estão, na verdade, dispersos ao longo do dia, ocorrendo pela manhã, tarde e noite, isso é comprovado pelo comprimento do vetor médio dado por 0,3302, indicando uma alta dispersão nos horários de ocorrência dos acidentes, com um desvio padrão circular de 5 horas e 41 minutos (1,485 radianos). Isso reforça a ideia de que os acidentes não estão todos concentrados em um único período, mas distribuídos ao longo de diferentes momentos do dia.

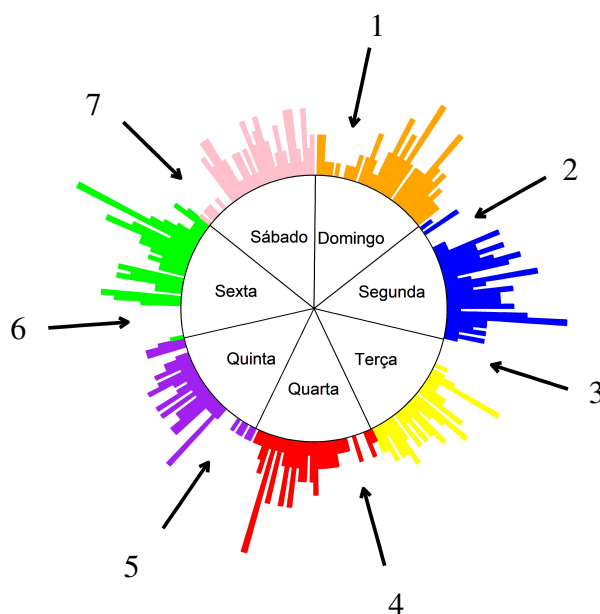
Esta alta dispersão pode ser ocasionada por influência do dia da semana em que ocorrem os acidentes, já que segundo Qu et al. (2021) nos dias úteis a taxa de acidentes em rodovias

é diferente daquelas ocorridas nos finais de semana. Dessa forma, a análise a seguir adota o agrupamento natural por dias da semana, substituindo a necessidade de gerar sete gráficos independentes por um único histograma circular permitindo evidenciar suas semelhanças e diferenças.

4.1 Método Empírico

Uma importante variável a ser estudada é o dia da semana em que ocorrem os acidentes. Na Figura 4.2, a divisão dos dias da semana é representada por sete segmentos, cada um correspondendo a uma mudança de dia, considerando início às 00h e término às 23h59min. Ao analisar as frequências dos acidentes de trânsito por dia da semana, adotou-se naturalmente uma distribuição 7-modal, com uma moda para cada dia da semana. As cores das barras representam os diferentes dias, no entanto nota-se que as falhas de ocorrências de acidentes pela madrugada observada na Figura 4.1 ficam mais evidentes em cada ciclo, o que pode indicar o final de cada um, que coincide com o início do próximo, conforme indicado pelas setas.

Figura 4.2 – Histograma circular dos acidentes ao longo da semana, com diferenciação de cores representando os dias da semana, conforme determinado pelo método empírico com horários limitantes, 00h as 23h59min.



Fonte: Do autor (2025).

Percebe-se ainda que as setas de 2 até 6 exibem uma falha de ocorrência mais acentuada do que as setas 7 e 1 que correspondem ao fim de semana em que, como esperado, há mais trânsito no período da madrugada. Isso de acordo com a afirmação feita por Wei et al. (2021).

Na Tabela 4.1, a quantidade de acidentes é informada de acordo com cada dia da semana. A coluna Limites tem o ângulo da meia noite de cada dia, considerando cada dia como $\frac{1}{7}$ dos 360° , assim cada dia corresponde a $51,43^\circ$ da circunferência, e por ultimo as colunas com os ângulos dos primeiros e últimos acidentes de cada dia de acordo com a delimitação empírica.

Tabela 4.1 – Número de acidentes (n), valores mínimos (min), máximos (max), limites (ou ponto de segmentação), média circular (μ), concentração (κ), arco entre os pontos mais próximos de agrupamentos vizinhos (*Arco*) e a proporção para cada dia determinado pelo Método Empírico.

Dia	n	min($^\circ$)	max($^\circ$)	Limites($^\circ$)	Arco($^\circ$)	μ($^\circ$)	κ
Domingo	94	0,96	50,64	51,43	2,18	30,42	20,38
Segunda	110	52,82	102,11	102,86	12,79	81,67	27,00
Terça	67	114,89	153,64	154,29	1,39	135,16	28,47
Quarta	77	155,04	203,82	205,71	2,21	184,25	24,17
Quinta	68	206,04	256,04	257,14	1,54	235,57	22,64
Sexta	94	257,57	307,61	308,57	1,25	288,96	29,32
Sábado	91	308,86	359,43	360,00	1,54	339,46	21,34

Fonte: Do autor (2025).

Nota-se que o maior valor de um dia praticamente coincide com o menor valor da classe seguinte, indicando uma continuidade na região e não um afastamento que caracterizaria uma mudança de ciclo ou agrupamento, que nesse caso serviria para evidenciar a mudança de dia da semana. Assim, as falhas indicadas não são contempladas por esses "limites empíricos".

Prossegue-se então com o teste para adequação de distribuições de Rao e Watson.

Na Tabela 4.1 foram testados os dias da semana individualmente, aqui as 24 horas do dia são distribuídos pelos 360° do círculo. No teste de Rao, somente na quinta feira não se rejeitou a hipótese dos acidentes estarem uniformemente distribuídos. Já no teste de Watson para uniformidade rejeitou-se a hipótese de uniformidade para todos os dias, individualmente ou agrupados.

Ao se testar, pelo teste de Watson, se os dias da semana seguem uma distribuição von Mises, rejeitou-se a hipótese nula para segunda, terça, sexta, sábado e dias comerciais. Assim,

Quadro 4.1 – Estatísticas dos Testes de Uniformidade e Distribuição von Mises para os Dias da Semana estabelecidos pela determinação convencional

Dia	Espaçamento de Rao	Watson	
		Uniforme	Von Mises
Domingo	Rejeita H_0	Rejeita H_0	Não Rejeita H_0
Segunda	Rejeita H_0	Rejeita H_0	Rejeita H_0
Terça	Rejeita H_0	Rejeita H_0	Rejeita H_0
Quarta	Rejeita H_0	Rejeita H_0	Não Rejeita H_0
Quinta	Não Rejeita H_0	Rejeita H_0	Não Rejeita H_0
Sexta	Rejeita H_0	Rejeita H_0	Rejeita H_0
Sábado	Rejeita H_0	Rejeita H_0	Rejeita H_0

Fonte: Do autor (2025).

percebe-se que há dificuldade pelo método empírico que considera o horário oficial de início e fim do dia como limites dos ciclos de horários de acidentes, já que no caso da uniforme, apenas a quinta feira seria ajustada e na von Mises três dias da semana (Domingo, Quarta e Quinta) seria ajustada a esta distribuição.

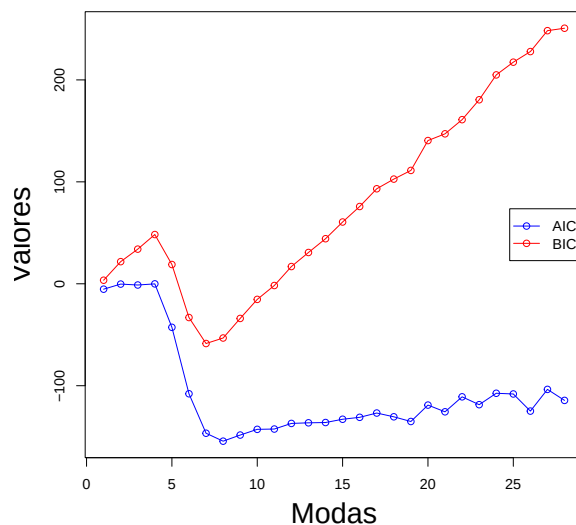
4.2 Método Probabilístico

Foi realizada a análise dos acidentes, ao longo da semana, utilizando modelos de misturas de von Mises, com o número de modas variando de 1 a 28, supondo até 4 modas por dia, sendo utilizado os critérios de informação de AIC e BIC para avaliar o número de componentes.

Conforme apresentado na Figura 4.3, o modelo ajustado para os acidentes considerando 1 a 5 modas apresentaram valores de AIC e BIC próximos de zero, sendo os mais simples e provavelmente insuficiente para capturar a complexidade dos dados.

Uma melhoria foi observada nos modelos considerando entre 6 e 8 modas, visto que com 7 modas o AIC foi de -146,6 e BIC de -58,6, indicando o melhor ajuste entre os modelos avaliados o que está de acordo com a expectativa dos sete dias da semana. Deve-se ressaltar que o AIC melhorou para o modelo com 8 modas, indicando uma possível bimodalidade em pelo menos um dia da semana, porém o BIC começou a penalizar esses modelos pela crescente complexidade. Nos modelos a partir de 9 modas, os valores de AIC e BIC aumentaram, à medida que se aumentava a quantidade de modas.

Figura 4.3 – Gráfico de linhas dos valores de AIC e BIC para até 28 modas dos modelos de misturas de von Mises



Fonte: Do autor (2025).

Assim o modelo com 7 modas, o de menor BIC, foi escolhido como o mais adequado para descrever a distribuição dos acidentes, ao longo da semana, sugerindo uma moda para cada dia da semana.

Vale ressaltar que o modelo de 8 modas, o de menor AIC, também é uma alternativa, mas, já que nesse modelo pelo menos um dos dias teria bimodalidade optou-se pelo modelo mais simples, respaldando-nos em estudos recentes que determinam a quantidade de modas em misturas de von Mises pelo BIC, Chatterjee e Ghosh (2024) estudou as mudanças direcionais exibidas pelas baleias azuis e jubartes caracterizada por uma mistura de nove distribuições von Mises, e em Jammalamadaka et al. (2021) que utilizaram o Critério de Informação Bayesiana (BIC) como método de estimativa de parâmetros para selecionar o número de componentes no modelo de mistura.

Considerando então as 7 modas, na primeira etapa do método probabilístico, tem-se que a expectativa, ou probabilidade de pertencer ao agrupamento será dado em função das proporções apresentadas na Tabela 4.2.

A Tabela 4.2 apresenta a distribuição dos acidentes, ao longo da semana, ordenada pela média circular (μ). Os valores de ω são baseados nos valores de n , ω representa a proporção de acidentes obtida pelo método de segmentação, enquanto n indica o número absoluto de acidentes observados em cada dia. Os dias com maior peso relativo no modelo são segunda-feira ($\omega = 0,178$) e sábado ($\omega = 0,170$), indicando que nesses dias ocorre uma quantidade

relativamente maior de acidentes em comparação aos outros dias da semana. Já o dia com menor proporção de acidentes foi quinta-feira ($\omega = 0,108$).

Tabela 4.2 – Número de acidentes (n), valores mínimos (min), máximos (max), proporção (ω_k), média circular (μ), concentração (κ), arco entre os pontos mais próximos de agrupamentos vizinhos ($Arco$) e a proporção para cada dia determinado pelo Método Probabilístico por algoritmo EM.

Dia	n	min(°)	max(°)	Arco(°)	μ(°)	κ	ω
Domingo	87	13,68	55,93	7,43	34,79	33,33	0,14
Segunda	106	63,36	102,11	12,78	82,41	30,60	0,18
Terça	73	114,89	162,93	3,78	137,01	20,62	0,12
Quarta	75	166,71	211,39	1,90	187,27	27,68	0,13
Quinta	65	213,29	257,57	14,32	237,33	27,36	0,11
Sexta	95	271,89	311,54	1,92	289,68	27,48	0,16
Sábado	100	313,46	372,43	1,25	343,09	15,25	0,17

Fonte: Do autor (2025).

Observa-se uma diferença importante, em relação ao método empírico, enquanto no método anterior, os menores valores de um dia praticamente coincidiam com os maiores valores da classe seguinte, sugerindo continuidade entre os dias da semana no presente método, os valores mínimos e máximos de cada dia estão bem mais afastados. Essa separação mais clara reflete uma segmentação mais robusta dos dias da semana, indicando que as regiões de vela estão efetivamente delimitando os diferentes ciclos diários.

Diferentemente do método empírico de análise, no qual todos os dias têm a mesma duração (iniciando e terminando à meia-noite), ao aplicar o agrupamento dos dias da semana para cada distribuição von Mises, as diferentes concentrações de cada dia resultam em ciclos diários com durações variadas, como mostra a Figura 4.4. Ressalta-se então, que neste método uma desvantagem é não ter um horário exato de limite entre os dias da semana.

Os dias com maior concentração de acidentes, ou seja, aqueles em que os acidentes tendem a ocorrer mais próximos dos horários de μ , são domingo ($\kappa = 33,333$) com os acidentes acontecendo por volta das 16h14min e segunda-feira ($\kappa = 30,6602$) com μ sendo 14h27min. Esses valores indicam que os acidentes nesses dias estão menos dispersos em torno da média, destacando um padrão mais consistente. Por outro lado, o dia com menor concentração de

Figura 4.4 – Histograma circular dos acidentes, ao longo da semana, com diferenciação de cores representando os dias da semana, conforme determinado pelo Método Probabilístico por algoritmo EM.



Fonte: Do autor (2025).

acidentes, ou seja, aquele em que os horários dos acidentes estão mais espalhados, em relação à média (μ), foi sábado ($\kappa = 15,250$).

Quadro 4.2 – Estatística dos Testes de Uniformidade e Distribuição von Mises para os Dias da Semana determinada pelo Método Probabilístico

Dia	Espaçamento de Rao	Watson	
		Uniforme	Von Mises
Domingo	Rejeita H_0	Rejeita H_0	Rejeita H_0
Segunda	Rejeita H_0	Rejeita H_0	Rejeita H_0
Terça	Não Rejeita H_0	Rejeita H_0	Não Rejeita H_0
Quarta	Não Rejeita H_0	Rejeita H_0	Não Rejeita H_0
Quinta	Rejeita H_0	Rejeita H_0	Não Rejeita H_0
Sexta	Rejeita H_0	Rejeita H_0	Rejeita H_0
Sábado	Não Rejeita H_0	Não Rejeita H_0	Não Rejeita H_0

Fonte: Do autor (2025).

Os resultados apresentados na Tabela 4.2 revelam maior quantidade de não rejeições ao teste de Rao se comparada ao método empírico, antes uma e agora três, e o teste de Watson para uniformidade sem rejeições anteriormente, não rejeita uniformidade para sábado, indicando que por este agrupamento alguns dias da semana, principalmente sábado, podem estar unifor-

memente distribuídos quando distribuídos. Isso é coerente com a alta dispersão observada para para este dia. Destaca-se também que o sábado não é considerado comercialmente um dia útil e desta forma é esperado que o tráfego nesse dia seja diferente dos demais.

Da mesma forma, poderia-se esperar que tal fato ocorresse no domingo, no entanto nesse dia não ocorre uniformidade e tem uma concentração maior de acidentes ($\kappa = 33,333$) e uma média circular de $\mu = 34,791$ (16h41) no fim do dia. Fato esse, coerente com acidentes de retorno de fim de semana.

Ao se analisar a adequação à distribuição von Mises, houve um aumento de não rejeições com relação ao método anterior, antes três dias não rejeitaram a hipótese nula (Domingo, Quarta e Quinta), enquanto pelo método probabilístico foram quatro: Terça, Quarta, Quinta e Sábado.

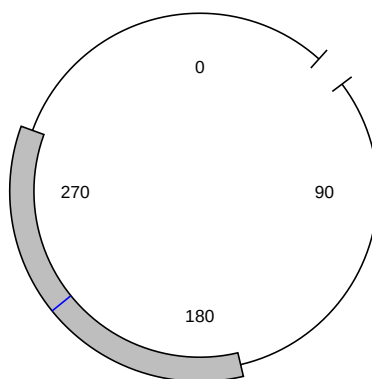
O método probabilístico, baseado em misturas de distribuições von Mises, demonstrou ser uma ferramenta mais robusta e informativa para análise da distribuição de acidentes ao longo da semana. Diferentemente do método empírico, que utiliza limites fixos baseados no horário oficial de início e fim do dia, o modelo probabilístico permite a adaptação às características intrínsecas dos dados, ajustando-se às concentrações e dispersões de acidentes em cada dia da semana. A segmentação mais clara entre os ciclos diários, evidenciada pelos limites mínimos e máximos mais bem definidos, reflete uma melhora ao se optar pelo uso do método probabilístico na identificação de padrões e na delimitação dos ciclos semanais.

4.3 Método de Segmentação por Boxplot Circular

Para realização do método do Boxplot Circular proposto neste trabalho, considera-se inicialmente simetria entre os dados 7-modais dias das semanas, ou seja, cada dia durando 24 horas. Multiplicou-se os valores por 7 obtendo dados unimodais, ou simplesmente horários dos acidentes nas rodovias mineiras, sem levar o dia da semana em consideração. Para determinar os limites dos ciclos diários de acidentes pelo Método do Boxplot Circular, é apresentado na Figura 4.5 o boxplot a partir dos dados unimodais como recomendado por Landler et al. (2018). Obteve-se o valor de 231° (15h24min) para a mediana circular, os quartis foram 290° (19h20min) e 167° (11h07min), enquanto os limites horário e anti-horário foram 53° (02h48min) e 42° (03h32min), respectivamente.

Os limites horário e anti-horário $L_H = 42^\circ$ e $L_A = 53^\circ$ foram determinados, e o ponto para separar os ciclos para os dados unimodais, horários dos acidentes durante o dia, será a

Figura 4.5 – Boxplot circular dos horários dos acidentes diários.



Fonte: Do autor (2025).

média circular de $\tilde{\theta}_L^* = 47,5^\circ$ (03h10min). Assim, nesse método, assume-se que o dia inicia-se às 03h10min e se encerra 24 horas depois. No método do boxplot circular, o primeiro ponto para separar os dias da semana nos dados 7-Modais, acidentes no decorrer da semana, foi $\frac{47,5}{7} = 6,79$. Para encontrar os outros 6 pontos adicionou-se $\frac{360}{7}$ para encontrar cada limite, resultando em 58,22; 109,65; 161,08; 212,50; 263,93; 315,36; e por ultimo 366,79 que completa o círculo em 6,79 como mostra a Tabela 4.3.

Tabela 4.3 – Número de acidentes (n), valores mínimos (min), máximos (max), arco entre os pontos mais próximos de agrupamentos vizinhos (*Arco*), média circular (μ) e concentração (κ) para cada dia determinado pelo Método Boxplot Circular.

Dia	n	min(°)	max(°)	Limites	Arco(°)	μ(°)	κ
Domingo	89	9,21	55,93	58,22	2,29	34,22	30,14
Segunda-feira	106	63,36	102,11	109,65	7,54	82,66	33,15
Terça-feira	71	114,89	157,71	161,08	3,37	136,37	24,72
Quarta-feira	77	162,89	211,39	212,50	1,11	186,91	27,11
Quinta-feira	65	213,29	257,57	263,93	6,36	237,53	29,31
Sexta-feira	97	271,89	314,14	315,36	1,22	290,21	27,67
Sábado	96	318,00	366,00	366,79	3,21	342,79	21,13

Fonte: Do autor (2025).

Enquanto no método empírico os menores valores de um dia praticamente coincidem com os maiores valores da classe seguinte, sugerindo continuidade entre os dias da semana, no método do Boxplot Circular os valores mínimos e máximos de cada dia estão bem mais afastados, assim como aconteceu com o método probabilístico.

Figura 4.6 – Histograma circular, dos acidentes ao longo da semana, com diferenciação de cores representando os dias da semana, conforme determinado pelo método boxplot.



Fonte: Do autor (2025).

A Figura 4.6 mostra essa segmentação mais evidente, a partir das falhas de ocorrência de acidentes entre os dias, adicionado ao fato dos limites que determinam o início e fim dos dias da semana está igualmente espaçados, resultando em uma segmentação mais robusta dos dias da semana, apresentando limites determinados pelo Boxplot Circular respeitando a duração de 24 horas de cada dia.

Percebe-se a uniformidade das cores, ao longo dos ciclos diários, sendo então uma adaptação do Método Empírico mais eficiente e simples que o método probabilístico, levando em consideração a distribuição dos horários dos acidentes em torno do círculo, o que resulta em um deslocamento do início e fim de cada ciclo.

A Tabela 4.3 compara os resultados dos testes de Rao e Watson em que, para o teste de Watson, rejeitou-se uniformidade para todos os dias da semana, diferente do teste de Rao que, assim como no método probabilístico, não rejeita uniformidade para terça, quarta e sábado. O teste de Watson para distribuição von Mises, teve a maior quantidade de não rejeição entre os métodos, sendo esse igual a cinco.

Permanece então, uma indefinição para a distribuição adequada para segunda e sexta. Observa-se que essa situação pode estar ligada ao fato de que nesses dias realmente é esperada uma mudança de hábitos de tráfego já que eles determinam a início e fim dos dias úteis.

Esse resultado ao ser comparado com os demais pode trazer à tona algumas discussões pois, conforme apresentado na Tabela 4.4, pode-se observar que o Método Empírico foi o menos

Quadro 4.3 – Estatística dos Testes de Uniformidade e Distribuição von Mises para os Dias da Semana determinado pelo Método do Boxplot Circular

Dia	Espaçamento de Rao	Watson	
		Uniforme	Von Mises
Domingo	Rejeita H_0	Rejeita H_0	Não Rejeita H_0
Segunda	Rejeita H_0	Rejeita H_0	Rejeita H_0
Terça	Não Rejeita H_0	Rejeita H_0	Não Rejeita H_0
Quarta	Não Rejeita H_0	Rejeita H_0	Não Rejeita H_0
Quinta	Rejeita H_0	Rejeita H_0	Não Rejeita H_0
Sexta	Rejeita H_0	Rejeita H_0	Rejeita H_0
Sábado	Não Rejeita H_0	Rejeita H_0	Não Rejeita H_0

Fonte: Do autor (2025).

Tabela 4.4 – Teste de Watson para adequação à distribuição von Mises dos acidentes distribuídos na semana nos três métodos estudados, em que R indica rejeição de H_0 e N a não rejeição, com a última coluna sendo a porcentagem de rejeições.

Critério	Dom	Seg	Ter	Qua	Qui	Sex	Sab	Total (%)
Empírico	N	R	R	N	N	R	R	57,14%
Probabilístico	R	R	N	N	N	R	N	42,86%
Boxplot Circular	N	R	N	N	N	R	N	28,57%

Fonte: Do autor (2025).

eficaz na avaliação da adequação dos acidentes distribuídos durante a semana, à distribuição von-Mises, já que obteve o maior número de rejeições (R), especificamente nas segundas, terças, sextas e sábados.

O Método Probabilístico por algoritmo EM, por outro lado, apresentou um desempenho melhor, com uma taxa de rejeição menor (42,86%) em comparação com o método empírico. Esse avanço deve-se ao fato de que o método probabilístico consegue ajustar melhor os horários de início e fim dos dias, permitindo uma análise mais refinada.

O Método de Segmentação por Boxplot Circular foi o mais eficaz, apresentando o menor percentual de rejeição (28,57%). Esse método conseguiu estimar dias com a mesma duração, como esperado, e indicou que a distribuição dos acidentes ao longo da semana ajusta-se melhor à distribuição von-Mises.

5 CONCLUSÕES

Ao analisar as ocorrências de acidentes, com base na divisão usual dos dias, a falta de limites definidos estatisticamente comprometeu a eficácia dos resultados, uma vez que o método empírico não utiliza modelos probabilísticos capazes de oferecer uma avaliação mais confiável. Com o Método Probabilístico por Algoritmo EM foi notória a melhora mas encontrou diferentes valores de concentração, o que não é desejado, já que se espera que os dias tenham a mesma duração, o que compromete a uniformidade necessária para uma análise mais precisa. Além disso, o esforço teórico e computacional exigido para esse método é maior, dado que são necessários ajustes mais complexos para alcançar os resultados obtidos. Pelo Método do Boxplot Circular, a sua capacidade de manter a uniformidade da duração dos dias e a alta taxa de não rejeição indicam que esse foi o método mais adequado para este estudo, ao equilibrar simplicidade e precisão de forma eficiente. Além disso, o Boxplot Circular se destaca por ter um baixo esforço computacional e ser teoricamente simples, o que o torna uma abordagem prática e eficaz para esse tipo de análise.

REFERÊNCIAS

- BRUZZONE, L.; PRIETO, D. F. An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images. **IEEE Transactions on image processing**, IEEE, v. 11, n. 4, p. 452–466, 2002.
- BUTTARAZZI, D. **Package ‘bpDir’**. [S.l.], 2022. R package version 1.0.0.
- BUTTARAZZI, D.; PANDOLFO, G.; PORZIO, G. C. **A boxplot for circular data**. [S.l.]: Oxford, 2018.
- CHASANI, P.; LIKAS, A. The uu-test for statistical modeling of unimodal data. **Pattern Recognition**, Elsevier, v. 122, p. 108272, 2022.
- CHASANI, P.; LIKAS, A. Statistical modeling of univariate multimodal data. **arXiv preprint arXiv:2412.15894**, 2024.
- CHATTERJEE, D.; GHOSH, P. On spatio-temporal directional association of blue and humpback whale’s migratory path navigation with sun, moon, ocean current & earth’s magnetic field. **ESS Open Archive eprints**, v. 88, p. 08896939, 2024.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. **Journal of the royal statistical society: series B (methodological)**, Wiley Online Library, v. 39, n. 1, p. 1–22, 1977.
- DESLANDES, S. F. O atendimento às vítimas de violência na emergência: "prevenção numa hora dessas?". **Ciência & Saúde Coletiva**, SciELO Public Health, v. 4, p. 81–94, 1999.
- HORNIK, K.; LEISCH, F. **movMF: Mixture of von Mises-Fisher Distributions**. [S.l.], 2014. R package version 1.2.0.
- JAMMALAMADAKA, S. R.; WAINWRIGHT, B.; JIN, Q. Functional clustering on a circle using von mises mixtures. **Journal of Statistical Theory and Practice**, Springer, v. 15, p. 1–17, 2021.
- KHARE, N. K. N. *et al.* Epidemiological study of road traffic accident cases attending tertiary care hospital in bhopal, madhya pradesh. 2012.
- LADEIRA, R. M.; BARRETO, S. M. Fatores associados ao uso de serviço de atenção pré-hospitalar por vítimas de acidentes de trânsito. **Cadernos de Saúde Pública**, SciELO Public Health, v. 24, p. 287–294, 2008.
- LANDLER, L.; RUXTON, G. D.; MALKEMPER, E. P. Circular data in biology: advice for effectively implementing statistical procedures. **Behavioral ecology and sociobiology**, Springer, v. 72, p. 1–10, 2018.
- LUND, U.; AGOSTINELLI, C.; AGOSTINELLI, M. C. **Package ‘circular’**. [S.l.], 2017. v. 775, n. 5.
- MARÍN, L.; QUEIROZ, M. S. A atualidade dos acidentes de trânsito na era da velocidade: uma visão geral. **Cadernos de Saúde Pública**, SciELO Public Health, v. 16, p. 7–21, 2000.

MEER, L. J. P. van der; B., K. M. **movMF: Mixture of von Mises-Fisher Distributions**. [S.l.], 2021. R package version 1.0.0.

MOONEY, J. A.; HELMS, P. J.; JOLLIFFE, I. T. Fitting mixtures of von mises distributions: a case study involving sudden infant death syndrome. **Computational Statistics & Data Analysis**, Elsevier, v. 41, n. 3-4, p. 505–513, 2003.

QU, X. *et al.* Exploring the influences of point-of-interest on traffic crashes during weekdays and weekends via multi-scale geographically weighted regression. **ISPRS International Journal of Geo-Information**, MDPI, v. 10, n. 11, p. 791, 2021.

SEYED, T. H. *et al.* Time analysis of fatal traffic accidents in fars province of iran. **Chinese journal of traumatology**, Chinese Medical Journals Publishing House Co., Ltd. 42 Dongsi Xidajie ..., v. 16, n. 02, p. 84–88, 2013.

TEAM, C. R. **R: A Language and Environment for Statistical Computing**. [S.l.], 2025. R Foundation for Statistical Computing, Vienna, Austria.

WEI, F. *et al.* Analysis of roadside accident severity on rural and urban roadways. **Intell. Autom. Soft Comput**, v. 28, p. 753–767, 2021.

WICKHAM, H. **ggplot2: Create Elegant Data Visualisations Using Grammar of Graphics**. [S.l.], 2016. R package version 3.4.0.

Artigo 3 - Elaborado conforme a norma NBR 6022	
Título do artigo:	Influência de fatores ambientais e sazonais em São João del Rei na captura de enxames de abelhas africanizadas
Autores:	Luiz Fernando Silva Resende Carla Regina Guimarães Brighenti Deodoro Magno Brighenti



**ARTIGO 3 – INFLUÊNCIA DE FATORES AMBIENTAIS E SAZONAIS EM SÃO
JOÃO DEL REI NA CAPTURA DE ENXAMES DE ABELHAS AFRICANIZADAS
INFLUENCE OF ENVIRONMENTAL AND SEASONAL FACTORS IN SÃO JOÃO
DEL REI ON THE CAPTURE OF AFRICANIZED BEE SWARMS¹**

Luiz Fernando Silva Resende

<https://orcid.org/0000-0003-4697-6503>

Carla Regina Guimarães Brighenti

<https://orcid.org/0000-0002-7822-3744>

Deodoro Magno Brighenti

<https://orcid.org/0000-0002-0612-1802>

Resumo: Analisou-se o sucesso da captura de enxames de abelhas africanizadas em ninhos-armadilha na Fazenda Experimental Risoleta Neves, São João del Rei, entre setembro de 2023 e agosto de 2024. O estudo investigou variáveis ambientais (precipitação, umidade, velocidade do vento, época do ano) e de configuração dos ninhos (cor, altura, ângulo do alvado), utilizando estatísticas lineares e circulares. Identificou-se bimodalidade nas épocas capturas, com picos no outono e primavera, associados a baixa precipitação e maior temperatura. Ninhos revestidos nas cores pretas e azuis, em alturas acima de 1,70 m, apresentaram maior sucesso. Concluiu-se que fatores ambientais e sazonais otimizam a captura, orientando práticas apícolas.

Palavras-chave: práticas apícolas; diagrama de rosas; histograma circular; boxplot circular; estatísticas circulares.

Abstract: The success of capturing Africanized bee swarms in trap nests at Risoleta Neves Experimental Farm, São João del Rei, from September 2023 to August 2024, was analyzed. The study investigated environmental variables (precipitation, humidity, wind speed, time of year) and nest configurations (color, height, entrance angle) using linear and circular statistics. A bimodal pattern was identified in capture periods, with peaks in autumn and spring, associated with low precipitation and higher temperatures. Black and blue nests, placed above 1.70 m, showed greater success. It was concluded that environmental and seasonal factors optimize swarm capture, guiding apicultural practices.

Keywords: apicultural practices; rose diagram; circular histogram; circular boxplot; circular statistics.

¹ Artigo científico apresentado à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária.

1 INTRODUÇÃO

A nidificação dos ninhos de abelhas ocorre de formas variadas, podendo ser em árvores, diretamente no solo ou até mesmo em rachaduras de estruturas construídas. Macedo *et al.* (2020) realizaram uma pesquisa com 60 ninhos de meliponínios, abelhas indígenas sem ferrão, constatando que 85% das abelhas nativas preferem cavidades em árvores vivas para nidificar, enquanto 6,67% optam por árvores mortas, 5% por troncos secos e 3,33% por cupinzeiros. Conforme Macedo *et al.* (2020), o entendimento sobre a nidificação das abelhas melíponas não se limita a encontrar uma cavidade adequada e uma florada favorável; aspectos como a orientação magnética dos ninhos também são relevantes. Esse fator tem intrigado estudiosos, que buscam compreender a direção para a qual os orifícios de entrada dos ninhos são posicionados.

Lacerda *et al.* (2017) e Macedo *et al.* (2020) realizaram estudos sobre a influência magnética, considerando os pontos cardeais e colaterais aos hábitos de nidificação de abelhas nativas. Em outro estudo, Vaz *et al.* (2021) determinaram que o ângulo de nidificação das abelhas *Melipona subnitida* foi de 2,37 graus, enquanto para *Frieseomelitta sp.* alcançou 45,68 graus.

Análises feitas com ressonância paramagnética eletrônica mostram a existência de material magnético (nanopartículas) de magnetitas no abdômen de abelhas *Apis mellifera* (Vaz *et al.*, 2021; Gould; Kirschvink; Deffeyes, 1978) e em abelhas sem ferrão *Schwarziana quadripunctata* (Lucano, 2006; Esquivel *et al.*, 2007). A existência de magnetita no abdômen das abelhas *A. mellifera* possibilita o desenvolvimento de modelos que explicam os mecanismos de orientação desses insetos. A capacidade de navegação das abelhas está relacionada a pistas visuais do ambiente, memorizadas por elas, que orientam seus deslocamentos entre o ninho e as áreas de forrageamento, conforme apontado por Capaldi e Dyer (1999).

Segundo Beier e Lindauer (1970), a posição do sol exerce um papel significativo nesse comportamento, combinada a marcos visíveis próximos ao ninho, os quais as abelhas podem armazenar na memória como um padrão característico, funcionando de maneira análoga a um ‘relógio solar’.

Ainda sobre o ambiente aos quais as abelhas fazem suas atividades, as diferentes condições climáticas interferem diretamente nas atividades das abelhas *A. mellifera*. Sendo a temperatura o principal fator associado ao aumento da produção de mel (Silva *et al.*, 2023), e a umidade, quando elevada, dificulta a coleta de recursos florais (Sousa; Rizzardo, 2014).

Além disso, mudanças climáticas, como as abordadas por Pinto *et al.* (2024), indicam que variações em temperatura e umidade podem alterar a fisiologia e o comportamento das abelhas africanizadas, reduzindo sua resistência a parasitas como o ácaro *Varroa destructor*.

Couvillon *et al.* (2014) demonstraram que abelhas africanizadas, precisam viajar significativamente mais longe no verão em comparação com a primavera ou o outono para coletar alimento, indicando que o verão representa as condições climáticas mais desafiadoras para encontrar recursos.

De igual modo, pode-se esperar que tal influência também ocorra no processo de migração de abelhas, no qual os indivíduos procuram novos locais para instalar seus enxames. Esta migração ocorre geralmente em períodos de escassez de alimento para buscar novos recursos alimentares ou no período de excesso devido à facilidade de multiplicação de enxames. Esse deslocamento está frequentemente associado ao período de enxameação, que, conforme (Leven *et al.*, 2006), tende a se concentrar na primavera ou no início da estação seca.

Durante esta migração, os apicultores instalam ninhos-armadilha para captura de enxames, que podem ser utilizadas na captura de abelhas com ou sem ferrão, sem trazer danos à vegetação nativa, bem como às espécies capturadas, pois as mesmas são alocadas em árvores sem a necessidade de intervenções, não agredindo o meio ambiente. Além do método de captura ser de baixo custo aos meliponicultores e apicultores, é também de fácil confecção (Vinícius-Silva; Ferraz; Werneck, 2021).

Existem modelos específicos preferenciais para cada espécie, no entanto, não é raro que um ninho-armadilha instalado para capturar um determinado tipo de abelha acabe resultando na captura de outra espécie. Vários são os fatores que podem influenciar no sucesso desta captura, tais como altura de instalação dos ninhos, locais como troncos secos, árvores e cupinzeiros, bem como seu posicionamento em relação ao sol e às condições climáticas locais (Qu *et al.*, 2022).

Steffan-Dewenter e Kuhn (2003) demonstraram que abelhas melíferas (*A. mellifera*) ajustam suas distâncias de forrageamento e atividades de dança em função da complexidade da paisagem, com forrageadoras de pólen percorrendo distâncias significativamente maiores em paisagens simples do que em paisagens complexas, especialmente em períodos de escassez de recursos. Reforçando a necessidade de considerar a heterogeneidade ambiental e a sazonalidade ao posicionar ninhos-armadilha, pois tais fatores podem afetar a atratividade e o sucesso na captura de enxames.

Nas épocas de enxameação (períodos naturais de divisão e deslocamento de enxames), é comum que os apicultores distribuam alguns ninhos perto de fontes de água ou de locais com boas floradas, por exemplo. A cada período, usualmente de 10 dias, é necessário observar se os ninhos foram povoadas, ou seja, se houve a captura (atração) de algum enxame. Se houver captura, o enxame deve ser transportado para o apiário ou meliponário em alguns dias (apenas o necessário para o início da postura pela rainha), pois, sem o acúmulo de alimento, o enxame será menos defensivo, e isso facilita seu transporte (Pereira *et al.*, 2016).

Entretanto, identificar as épocas de enxameação para realizar as instalações dos ninhos-armadilha pode ser um desafio, já que esse processo depende da verificação de padrões que favorecem o sucesso da captura de enxames considerando variáveis como altura, cor de revestimento e posicionamento angular das entradas dos ninhos, bem como condições climáticas, incluindo precipitação, umidade, velocidade e direção do vento. Que podem otimizar atividade de apicultores e meliponicultores.

Ressalta-se que muitas dessas variáveis têm dependência com a época do ano, e assim pode ocasionar eventos de captura com frequência de maior relevância em mais de um período ao longo do ano, caracterizando um conjunto de dados chamado multimodal.

Assim, o estudo busca caracterizar o sucesso de captura de abelhas africanizadas em função de fatores que influenciam a atratividade dos ninhos-armadilha verificando a ocorrência de multimodalidade.

As variáveis analisadas no estudo abrangem tanto características qualitativas (Cor do revestimento do ninho) quanto quantitativas, sendo estas últimas subdivididas em dois tipos: lineares e circulares. As variáveis lineares incluem altura, precipitação, umidade e velocidade do vento, enquanto as circulares englobam o posicionamento angular das entradas dos ninhos e a direção do vento. Para explorar as relações entre essas variáveis, são empregadas técnicas de correlação adaptadas à sua natureza: correlações linear-linear, que avaliam associações entre variáveis lineares; linear-circular, que investigam interações entre variáveis lineares e circulares; e circular-circular, que examinam relações entre variáveis circulares, como ângulos de entrada do ninho armadilha, direção do vento e época de captura.

Além disso, ferramentas gráficas como boxplots circulares e diagramas de rosas são utilizadas para visualizar a distribuição de variáveis circulares, destacando *outliers* de maneira intuitiva e tendências sazonais para caracterização da multimodalidade das capturas de enxames de abelhas africanizadas. A abordagem multimodal do estudo, ao integrar essas técnicas

estatísticas e gráficas, permite capturar a complexidade das interações entre fatores lineares e angulares. Essas análises combinadas fortalecem a investigação, proporcionando uma base sólida para recomendações práticas e direcionadas aos produtores apícolas.

2 METODOLOGIA

A metodologia adotada neste estudo envolveu a caracterização detalhada da área experimental, a construção e posicionamento estratégico de ninhos-armadilhas, a sistematização da coleta de dados e a aplicação de análises estatísticas para avaliar variáveis ambientais e de captura.

2.1 Caracterização da área de instalação dos ninhos-armadilhas

A pesquisa foi desenvolvida na Fazenda Experimental Risoleta Neves, no *Campus* Tancredo de Almeida Neves (CTAN) da Universidade Federal de São João del-Rei (UFSJ). A cidade de São João del Rei, localizada no estado de Minas Gerais, Brasil, abrange uma área de 1.465,78 km² e está situada a uma altitude média de 860 metros. Seu clima é classificado como Cwa, segundo a classificação de Köppen, caracterizado por um inverno seco e um verão chuvoso, com temperatura média anual de 19,2 °C, temperatura mínima mensal de 15,5 °C e precipitação anual de 1437 mm. A região apresenta uma deficiência hídrica de 61 mm, ou seja, a quantidade de água disponível inferior à demanda hídrica das plantas, e aproximadamente quatro meses de estiagem (BARUQUI *et al.*, 2006).

O relevo da região é caracterizado por compartimentos planálticos dissecados, com altitudes variando entre 400 e 1.300 metros, predominando formas de colinas e morros de vertentes que configuram a típica paisagem de “mares de morros”. Os vales em “V” encaixados e a topografia fortemente ondulada a montanhosa são traços marcantes desse ambiente. A vegetação nativa é diversificada, incluindo formações como florestas tropicais (Netto *et al.*, 2023).

2.2 Montagem dos ninhos-armadilhas

Os ninhos-armadilhas foram confeccionados utilizando caixas de papelão do tipo arquivo (Figura 2.1 (a)), adquiridas com recursos da Universidade Federal de São João del-Rei (UFSJ). Essas caixas foram selecionadas pelo seu baixo custo e facilidade na montagem. O processo de montagem envolveu etapas específicas para garantir a funcionalidade e a durabilidade dos ninhos, considerando as condições ambientais da área experimental.

Figura 2.1 – Etapas da montagem dos ninhos-armadilhas: (a) caixa de papelão tipo arquivo, (b) adição do beiral fixado com fita adesiva, (c) aplicação de cera de abelhas na tampa interna e (d) revestimento com lona plástica impermeável.



Fonte: Do autor (2025).

Inicialmente, as caixas de papelão foram preparadas com a adição de um beiral na entrada (Figura 2.1 (b)), construído de papelão e fixado com fita adesiva branca. O beiral foi projetado para proteger o interior do ninho contra a entrada de chuvas, minimizando a deterioração do material e proporcionando um ambiente mais seguro para os enxames.

Na parte interna da tampa de cada caixa, foi aplicada cera de abelhas (Figura 2.1 (c)), espalhada de maneira uniforme. Essa etapa teve como objetivo atrair os enxames, uma vez que a cera emite odores familiares às abelhas, aumentando a probabilidade de ocupação dos ninhos-armadilhas. A cera utilizada foi obtida de fontes apícolas locais.

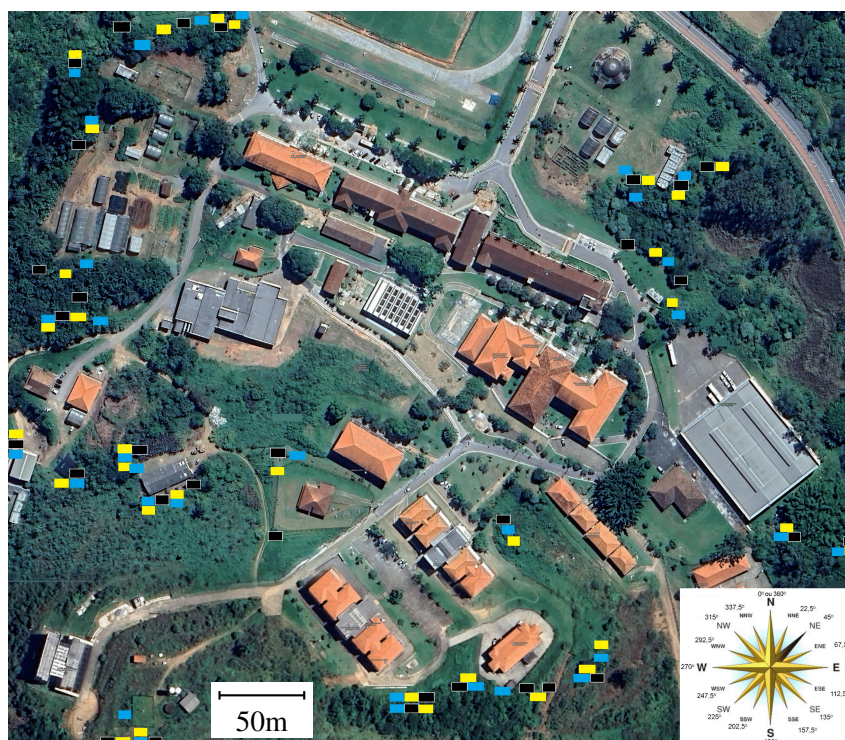
Para proteger as caixas contra intempéries e prolongar sua vida útil, cada ninho-armadilha foi revestido externamente com lona plástica de espessura de 2 mm, com proteção contra raios ultravioleta (UV) e característica impermeável (Figura 2.1 (d)). As lonas foram escolhidas nas cores amarela, azul e preta, permitindo a análise da influência da cor na atratividade dos ninhos. O revestimento foi fixado de forma a cobrir completamente a caixa, garantindo vedação contra umidade e resistência às condições climáticas da região.

2.3 Instalação dos ninhos-armadilhas

Foram instaladas 93 ninhos-armadilhas em uma área total de $0,162 \text{ km}^2$ situado no *Campus* CTAN da UFSJ, sendo 31 de cada cor, e distribuídos em trios em 31 locais, indicado no mapa da Figura 2.2.

Os ninhos foram numerados e posicionados em alturas variadas, sendo estas medidas com o auxílio de uma fita métrica em relação ao solo. A direção do alvado, bem como a

Figura 2.2 – Mapa dos ninhos-armadilhas instalados.



Fonte: Adaptado de Google Maps (2025)

altitude, latitude e longitude de cada ninho-armadilha, foi determinada com precisão utilizando o aplicativo *Bússola Digital Simples +Smart* (versão 3.1.0, desenvolvedor: Battery Stats Saver), disponível na Google Play Store, acessado em dispositivo Android.

Em cada captura realizada, o ninho-armadilha foi substituído por outro com as mesmas características para possibilitar recapturas, e os enxames capturados foram transferidos para o apiário ou meliponário didático da UFSJ, conforme a espécie, sendo o novo ninho-armadilha codificado para registro de repetições de captura.

Realizaram-se visitas periódicas, conforme condições climáticas, com o apoio do Grupo de Estudos em Apicultura (GEAPIS) da UFSJ, registrando-se possíveis capturas, a data de cada evento e o tempo decorrido entre a instalação e a captura, e os dados climáticos da data da observação. As observações ocorreram entre 1 de setembro de 2023 e 31 agosto de 2024.

2.4 Análise descritiva das variáveis ambientais e de captura

Os dados coletados para a análise descritiva das variáveis ambientais e de captura foram organizados em três grupos distintos:

- a) ninhos-armadilhas instalados (Geral): corresponde ao conjunto completo de ninhos posicionados, independentemente do sucesso na captura;
- b) ninhos-armadilhas sem captura de enxames (Ausente): inclui apenas os ninhos que não atraíram enxames durante o período de observação;
- c) ninhos-armadilhas com captura de enxames (Presente): abrange os ninhos que registraram pelo menos uma captura, considerando a reposição após cada evento.

Para cada ninho-armadilha, foram analisadas as seguintes variáveis:

- a) altura de instalação dos ninhos-armadilhas (Altura): medida em metros;
- b) cor do revestimento externo dos ninhos-armadilhas (Cor): classificada qualitativamente;
- c) orientação da entrada do ninho-armadilha (Ângulo do alvado): expressa em graus.

Para o último grupo, em que foram registrados uma ou mais capturas, devido à reposição dos ninhos, foram estudadas também as variáveis:

- a) época do ano da captura do enxame (Época): registrada em graus correspondentes ao dia do ano;
- b) precipitação total acumulada (Chuva): em milímetros;
- c) umidade relativa média (Umidade): em porcentagem;
- d) velocidade média do vento (Velocidade): em metros por segundo;
- e) direção média do vento (Direção): em graus.

As análises descritivas foram adaptadas à natureza de cada variável, considerando se eram qualitativas (Cor), quantitativas lineares (Altura, Chuva, Umidade, Velocidade) ou quantitativas circulares (Ângulo do alvado, Época, Direção).

2.5 Análise de correlação entre variáveis

Explorou-se a correlação entre diferentes tipos de variáveis quantitativas para compreender as relações entre os fatores ambientais e de configuração dos ninhos-armadilhas, para variáveis quantitativas lineares foi calculado o coeficiente de correlação de Pearson, para variáveis lineares e circulares utilizou-se a correlação circular-linear, e para pares de variáveis circulares aplicou-se a correlação circular de Fisher.

A correlação linear-linear (LL) é aplicada quando ambas as variáveis são quantitativas e pertencem ao domínio da reta real. Para avaliar a associação linear entre duas variáveis X e Y , com médias amostrais $\bar{x}$ e $\bar{y}$, o coeficiente de correlação de Pearson r_{xy} é definido como:

$$LL = r(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

no qual x_i e y_i representam os valores observados, e $\bar{x}$ e $\bar{y}$ são as médias amostrais respectivas. Essa medida quantifica o grau de associação linear entre variáveis e, para inferências estatísticas, idealmente requer uma distribuição aproximadamente normal. É amplamente aplicada em estudos com variáveis contínuas no conjunto dos números reais (Asuero; Sayago; González, 2006).

A correlação linear-circular (LC) é empregada quando uma variável é linear, situada na reta real, e a outra é circular, representada como um ângulo em um círculo unitário. Uma medida de correlação entre uma variável linear x e uma variável angular θ é a correlação linear-circular (Mardia, 1976). Considere uma amostra dada por $(x_1, \theta_1), \dots, (x_n, \theta_n)$, na qual x_i é a variável linear e θ_i é circular; a correlação é dada por:

$$LC = r(X, \theta) = \frac{r_{xc}^2 + r_{xs}^2 - 2r_{xc}r_{xs}r_{cs}}{1 - r_{cs}^2},$$

em que:

- a) $r_{xc} = r(x_i, \cos \theta_i)$;
- b) $r_{xs} = r(x_i, \sin \theta_i)$;
- c) $r_{cs} = r(\cos \theta_i, \sin \theta_i)$.

A correlação circular-circular (CC) é utilizada quando ambas as variáveis são angulares, localizadas em um círculo unitário, como a observação angular θ e a direção angular ϕ , com médias direcionais $\bar{\theta}$ e $\bar{\phi}$. Um coeficiente de correlação CC é calculado como:

$$CC = r(\theta, \phi) = \frac{\sum_{i=1}^n [\sin(\theta_i - \bar{\theta}) \sin(\phi_i - \bar{\phi})]}{\sqrt{\sum_{i=1}^n \sin^2(\theta_i - \bar{\theta}) \sum_{i=1}^n \sin^2(\phi_i - \bar{\phi})}},$$

no qual θ_i e ϕ_i são as variáveis circulares, e $\bar{\theta}$ e $\bar{\phi}$ são suas médias direcionais, obtidas por métodos de estatística circular (Jammalamadaka; Sengupta, 2001).

Para a variável qualitativa cor de revestimento, utilizou-se o coeficiente V de Cramer, calculado com base na estatística do teste Qui-Quadrado de Pearson χ^2 , no total de observações

n , no número de linhas m da tabela de contingência, e no menor valor entre os graus de liberdade $\min(m-1, n-1)$ (Santos; Carneiro; Oliveira, 2016), dado por:

$$V = \sqrt{\frac{\chi^2}{n \cdot \min(m-1, n-1)}},$$

em que N é o total de observações na amostra, m é o número de linhas da tabela de contingência, n é o número de colunas da tabela de contingência e $\min(m-1, n-1)$ é o menor valor entre os graus de liberdade das linhas ($m-1$) e das colunas ($n-1$). O coeficiente V varia entre 0 (nenhuma associação) e 1 (associação perfeita), indicando a força da relação entre as variáveis qualitativas analisadas.

3 RESULTADOS E DISCUSSÕES

Os Resultados e Discussões apresenta uma análise detalhada das capturas de enxames de abelhas africanizadas em ninhos-armadilhas instalados na Fazenda Experimental Risoleta Neves, no *Campus* Tancredo de Almeida Neves (CTAN) da UFSJ.

3.1 Mapa da captura dos ninhos-armadilhas instalados

O *Campus* Tancredo de Almeida Neves (CTAN), onde foi conduzido este estudo, está localizado em um declive de uma região montanhosa, apresentando uma variação topográfica relevante. A Figura 3.1 exibe um mapa de calor da altitude da área experimental, ilustrando claramente essa variação altitudinal que influencia as condições ambientais e, potencialmente, a distribuição dos ninhos-armadilhas.

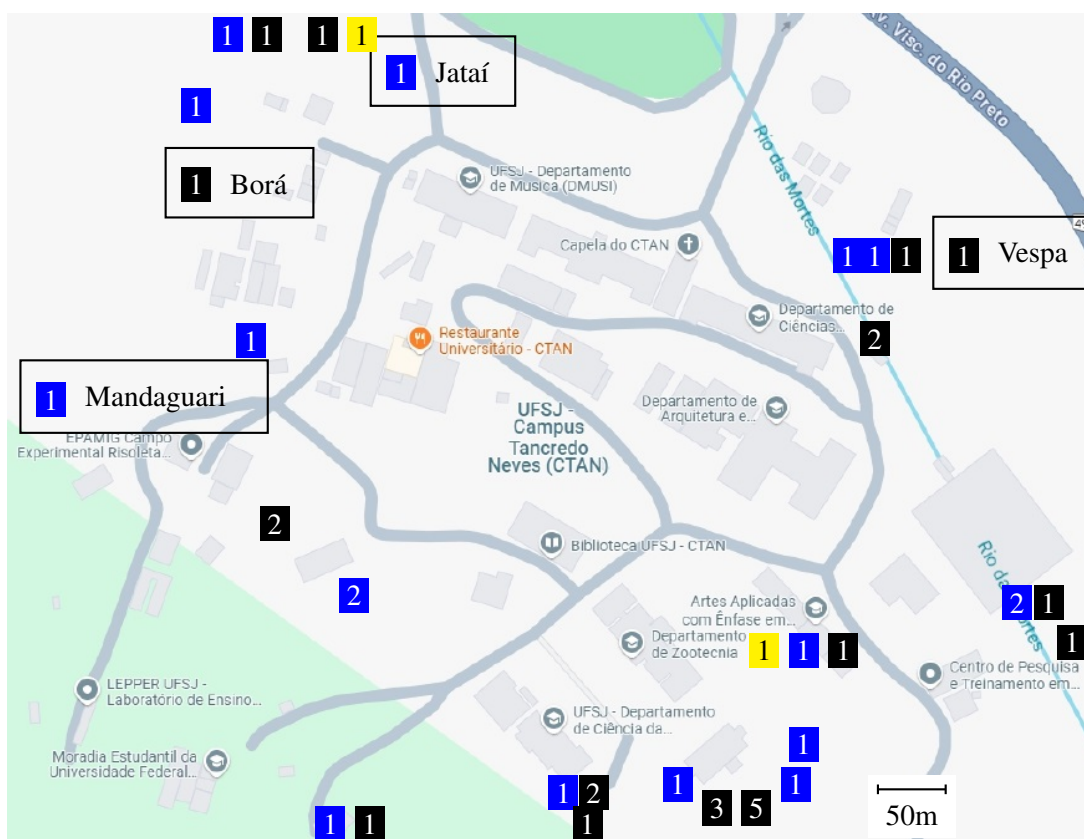
Figura 3.1 – Mapa de calor da altitude do *Campus* Tancredo de Almeida Neves (CTAN), ilustrando a variação topográfica da área experimental.



Fonte: Topographic Map (2025)

A Figura 3.2 ilustra a distribuição geográfica dos ninhos-armadilhas em que houve capturas de enxames na área experimental, com pontos representando os locais das capturas, identificados por cores conforme o revestimento (azul, preto ou amarelo) e rotulando capturas repetidas no mesmo local.

Figura 3.2 – Mapa dos ninhos-armadilha com capturas.



Fonte: Do autor (2025).

Em quatro ninhos-armadilhas foram capturados enxames das espécies *Tetragona clavipes* (Borá), *Tetragonisca angustula* (Jataí), *Polybia occidentalis* (Vespa) e *Scaptotrigona postica* (Mandaguari). Devido às especificidades dessas espécies, essas capturas não serão consideradas no estudo a seguir, somente os ninhos-armadilhas que capturaram enxames de abelhas africanizadas.

O mapa revela um total de 28 locais distintos onde ocorreram capturas de enxames, todos de abelhas africanizadas. Alguns locais exibiram múltiplas capturas, conforme indicado por valores superiores a 1, sugerindo que certas posições podem apresentar características ambientais ou posicionais favoráveis, a serem investigadas posteriormente.

Notavelmente, os ninhos-armadilha com maior número de repetições, como aqueles com três e cinco capturas, concentram-se em posições mais altas, conforme indicado pela análise de altitude no estudo. Essas posições correspondem a latitudes menores e longitudes que se alinham com a topografia elevada da Fazenda Experimental Risoleta Neves, sugerindo que a combinação de latitude e longitude pode favorecer a atratividade dos ninhos-armadilhas devido a fatores como exposição ao vento ou microclima local.

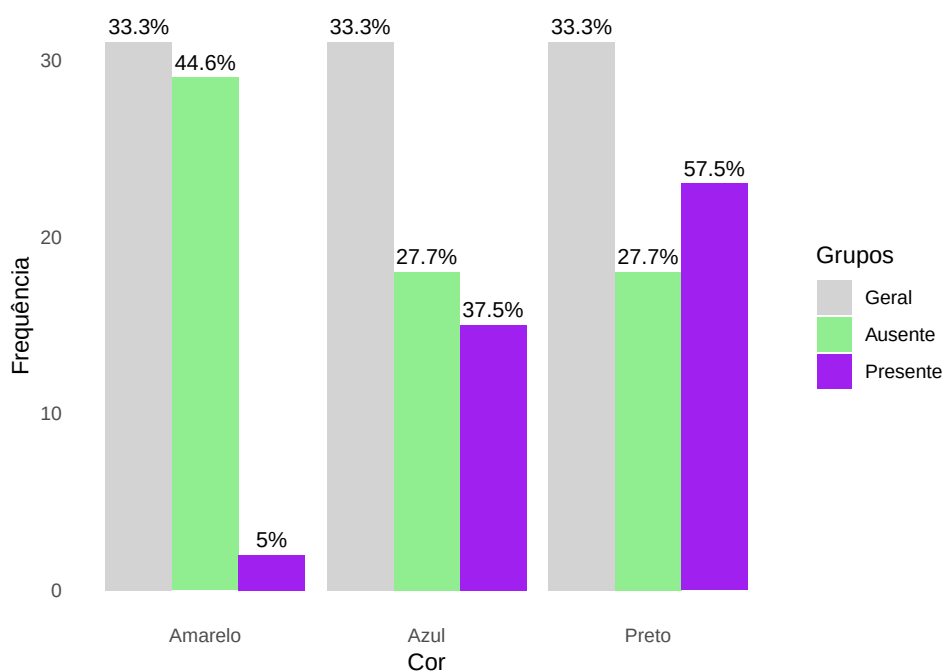
3.2 Estatística descritiva

Esta seção apresenta a análise descritiva das variáveis associadas aos ninhos-armadilhas instalados e às capturas de enxames realizadas na área experimental. Foram registradas 40 capturas nos 28 locais distintos. Os resultados são organizados em subseções que exploram cada variável separadamente e em pares, utilizando estatísticas descritivas lineares e circulares, bem como representações gráficas, para oferecer uma visão abrangente das interações entre as condições ambientais e o desempenho dos ninhos-armadilhas.

3.2.1 Cor de revestimento dos ninhos-armadilhas

A Figura 3.3 analisa a influência da cor de revestimento nos grupos Geral, Ausente e Presente. Os ninhos-armadilhas foram distribuídos em proporções iguais, com 31 (33,3%) para cada cor.

Figura 3.3 – Distribuição de cores de revestimento dos ninhos nos grupos: Instalados (Geral), Ausente e Presente.



Fonte: Do autor (2025).

Entre os ninhos com revestimento de cor amarela destacou-se com 29 ninhos das abelhas africanizadas a ausência de capturas. E por outro lado, para os ninhos azuis e pretos somente

18, sugerindo que os ninhos revestidos de azul e preto têm maior propensão a atrair enxames, enquanto os amarelos estão mais associados à ausência de capturas.

Nos ninhos com captura, complementando a análise anterior, a frequência foi menor para a cor amarela, com apenas 2 capturas. Porém diferente dos valores iguais para captura Ausente para as cores azul e preto, elas apresentaram diferença com 15 e 23 capturas, respectivamente, indicando que a recaptura foi maior nos ninhos-armadilha pretos (Tabela 3.1).

Tabela 3.1 – Distribuição das capturas de enxames por cor de revestimento dos ninhos-armadilhas.

Cor	Ausente	Presente	Total de Ninhos instalados com reposições
Amarelo	29	2	31
Azul	18	15	33
Preto	18	23	41

Fonte: Do autor (2025).

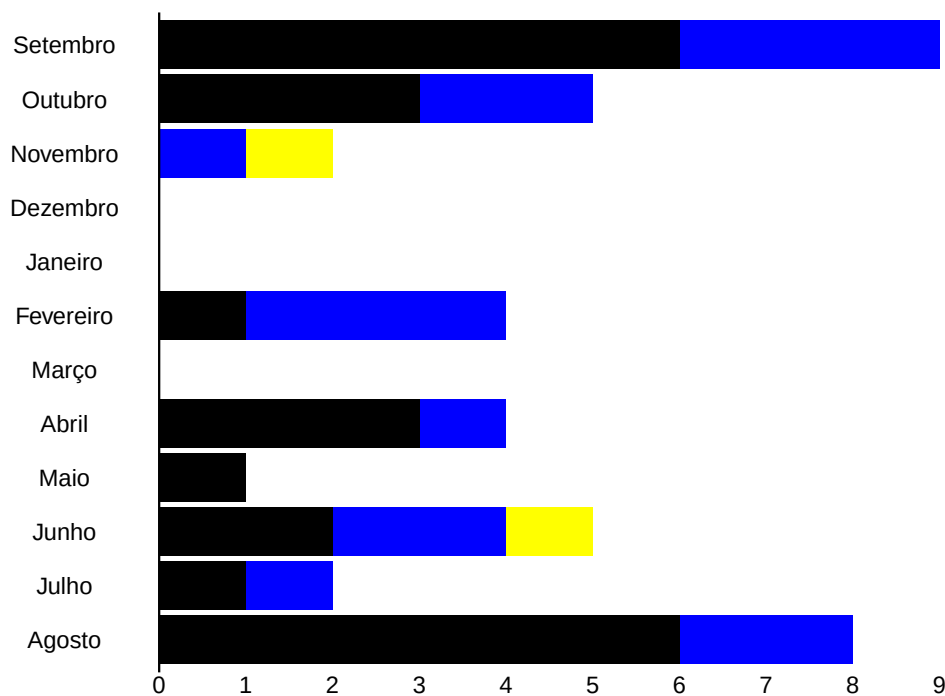
Para avaliar a associação entre as cores de revestimento dos ninhos-armadilha e o sucesso de captura, calculou-se o coeficiente V de Cramer, obtendo-se $V = 0,4315$, indicando uma associação moderada entre as variáveis categóricas. Além disso, a significância dessa associação foi verificada por meio do teste qui-quadrado, cuja estatística foi $\chi^2 = 19,55478$ com $p\text{-valor} < 0,05$, confirmando que há uma correlação significativa entre as cores de revestimento e o sucesso de captura.

3.2.2 Capturas por mês

A Figura 3.4 apresenta a distribuição das capturas de enxames de abelhas africanizadas ao longo dos meses, organizadas por cor de revestimento dos ninhos-armadilhas, em um período que se estende de setembro de 2023 a agosto de 2024.

Não houve registros de capturas em dezembro e janeiro, possivelmente devido ao período de chuvas intensas característico do clima Cwa da região, o que sugere uma sazonalidade nas capturas. Os meses com maior número de capturas foram agosto (8), setembro (9), outubro (5) e junho (5), indicando um padrão de atividade intensa durante a transição para a primavera, quando as condições climáticas e a disponibilidade de flores tendem a favorecer o forrageamento das abelhas.

Figura 3.4 – Distribuição das capturas por mês e cor dos ninhos-armadilhas, de setembro de 2023 a agosto de 2024.

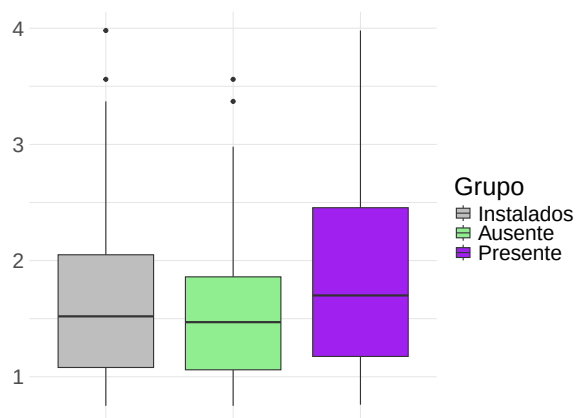


Fonte: Do autor (2025).

3.2.3 Altura de instalação dos ninhos-armadilhas

A altura de instalação dos ninhos-armadilhas foi avaliada para o sucesso da captura de enxames de abelhas africanizadas. As estatísticas descritivas da altura foram calculadas para a construção do boxplot, conforme apresentado na Figura 3.5.

Figura 3.5 – Boxplot comparando a altura entre os grupos Instalados, Presente e Ausente, sem distinção por cor de revestimento dos ninhos-armadilha.



Fonte: Do autor (2025).

Os boxplots da altura dos ninhos-armadilhas não revelam diferenças acentuadas entre os grupos analisados inicialmente. O grupo de ninhos Instalados possui mediana de 1,52 m, com o primeiro quartil em 1,08 m e o terceiro quartil em 2,05 m, em 93 observações. Os ninhos sem captura, pertencentes ao grupo Ausente, apresentam uma distribuição de alturas ligeiramente inferior, com mediana de 1,47 m, variando do primeiro quartil de 1,06 m ao terceiro quartil de 1,86 m, e uma média de 1,60 m com base em 65 observações.

Os ninhos do grupo de Captura Presente, destacam-se por apresentarem alturas superiores aos outros dois grupos, com mediana de 1,70 m com o primeiro quartil em 1,17 m e o terceiro quartil em 2,46 m, sendo a amplitude de 0,76 m a 3,98 m nas 40 observações. Pode-se investigar uma possível preferência dos enxames por posições mais altas.

Ao retirar os ninhos com captura do conjunto total de ninhos instalados, observa-se uma leve diminuição na média de altura, que passa de 1,68 m para 1,60 m. Indicando que os ninhos com captura tendem a estar posicionadas em alturas ligeiramente maiores.

O grupo Presente exibe uma mediana de 1,70 m, em 40 observações. Essa diferença sugere que as recapturas ocorreram predominantemente em ninhos posicionados a alturas mais elevadas, o que pode indicar uma preferência em ninhos-armadilhas instalados em locais mais altos.

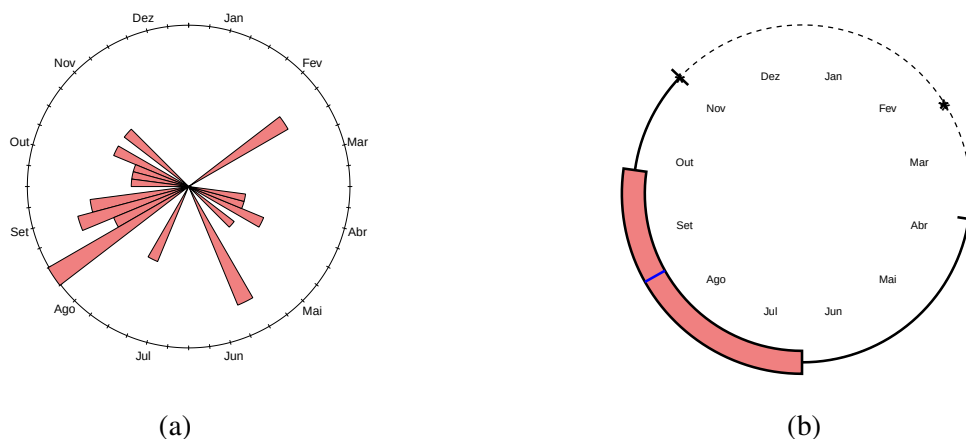
3.2.4 Época do ano

A análise da época do ano em que ocorreram as capturas de enxames de abelhas africanizadas foi realizada considerando o período de observação entre setembro de 2023 e agosto de 2024. Para isso, os dados foram convertidos em uma escala circular, representando os dias do ano em graus, na qual cada dia foi associado a um ângulo proporcional ao ciclo anual de 365 dias. Essa abordagem permite visualizar padrões sazonais de forma contínua.

A Figura 3.6(a) apresenta um diagrama de rosas que ilustra a distribuição das capturas ao longo do ano em radianos, destacando que as capturas não ocorreram de maneira uniforme, nem se concentrando em um único período específico. Já a Figura 3.6(b) exibe um boxplot circular incluindo 8 meses do ano, enquanto nos meses restantes, as capturas ocorridas foram classificadas como *outliers*.

Comparando as Figuras 3.6 (a) e 3.6 (b), observa-se nos meses entre novembro e fevereiro no diagrama de rosas há um trecho sem captura e o período de maio a agosto também

Figura 3.6 – (a) Diagrama de rosas da época das capturas (em radianos) de enxames de abelhas africanizadas, mostrando a distribuição circular ao longo dos 14 meses. (b) Box-plot circular da época das capturas (em graus), evidenciando a distribuição sazonal para dois grupos distintos.



Fonte: Do autor (2025).

tem baixa ocorrência de captura. No entanto, no boxplot circular da Figura 3.6 (b), os meses de julho e agosto situam-se na região interquartílica, apesar de registrarem apenas duas capturas. Isto revela que há um problema na representação do boxplot que não apresenta a realidade exibida na Figura 3.6 (a).

Diante do problema apresentado, buscou-se segmentar as épocas de captura de acordo com os métodos descritos no Capítulo 1, gerando um boxplot bimodal (Figura 3.7).

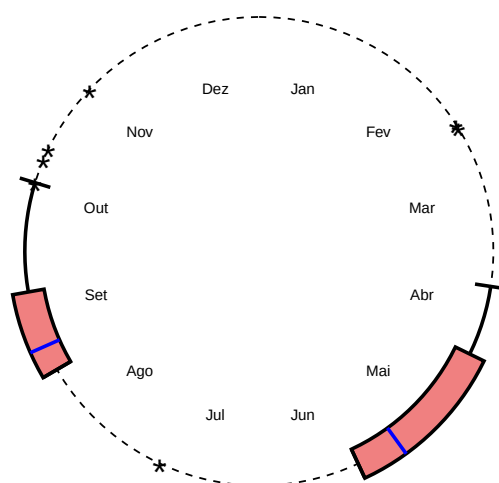
Para a caracterização do Método Empírico, as capturas foram segmentadas em primeiro e segundo semestre. Os Métodos Probabilísticos e o Boxplot Circular apresentaram resultados semelhantes, diferindo apenas nas capturas de julho, que não influenciaram a construção dos boxplots, pois foram posteriormente classificadas como *outliers* nos três métodos.

As capturas concentram-se majoritariamente em duas épocas distintas: uma entre 10 de abril e 06 de junho (outono) e outra entre 29 de agosto e 19 de outubro (primavera). Esses períodos têm condições climáticas específicas que serão discutidas a seguir.

3.2.4.1 Condições climáticas mensais lineares

A relação entre as capturas de enxames de abelhas africanizadas e as variáveis ambientais ao longo dos meses é ilustrada no diagrama de rosa da Figura 3.8, com valores normalizados no intervalo [0, 1]. Os valores de mínimo e máximo para escala de cada variável estão dispo-

Figura 3.7 – Boxplot circular bimodal mostrando a distribuição das capturas de enxames de abelhas africanizadas ao longo do ano (em graus) e eventos isolados representados por asteriscos.



Fonte: Do autor (2025).

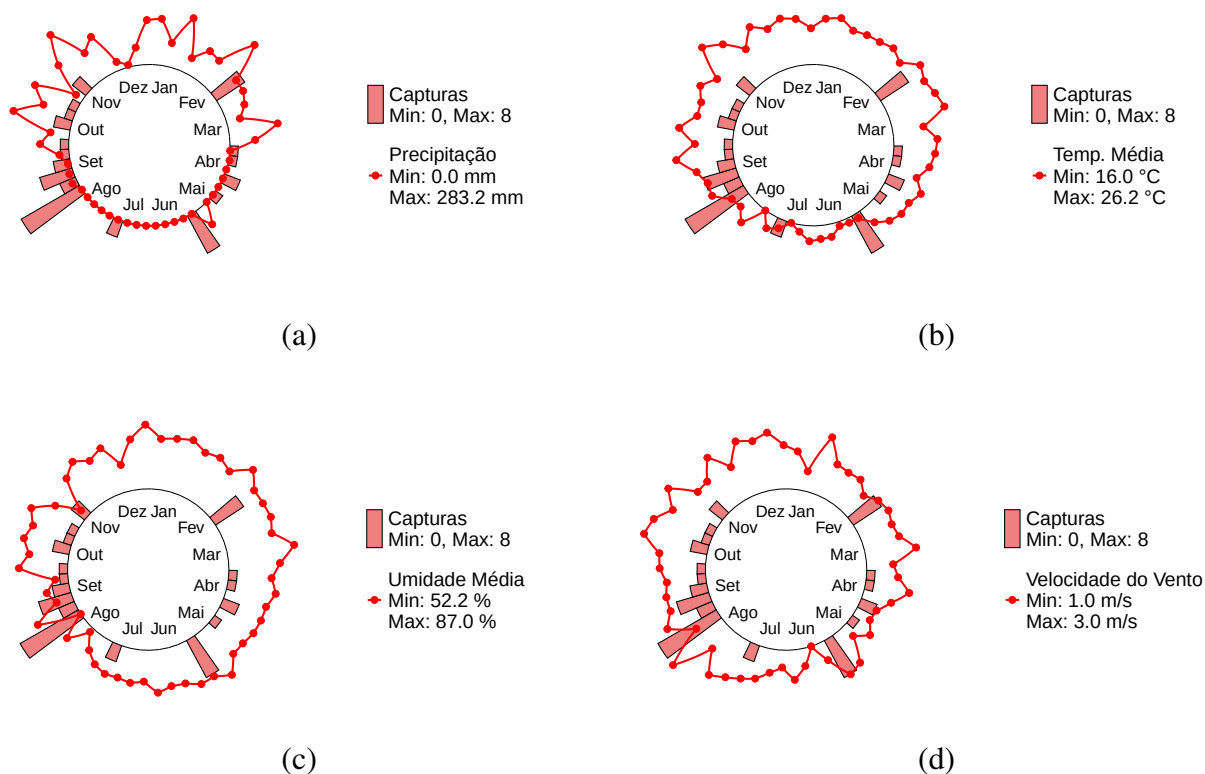
níveis na legenda. As capturas são exibidas como barras normalizadas enquanto as variáveis ambientais são representadas por linhas e pontos.

Pela Figura 3.8 (a), percebe-se que os meses com precipitações próximas a zero concentram os principais meses de capturas: abril (4), junho (5), agosto (8) e setembro (9), enquanto outubro, com uma maior precipitação, registra 5 capturas, pode ser explicado por uma transição para condições menos favoráveis. Esse período seco de abril a setembro está incluído nos limites do boxplot da Figura 3.6, indicando que a baixa precipitação tem uma relação positiva com as capturas de enxames de abelhas africanizadas. Já os períodos com os maiores níveis de precipitação são os meses que estão fora dos limites do boxplot, na região dos *outliers*.

A Figura 3.8 (b) revela diferenças em relação à temperatura média entre os dois períodos de maior atividade de enxameação de abelhas africanizadas determinados nos boxplots da Figura 3.7. Setembro e outubro são os períodos mais quentes do ano e concentram a maioria das capturas, os meses de junho, julho e agosto, caracterizados por temperaturas mais frias, formam uma região de separação entre os dois períodos de captura, já que as capturas registradas em junho ocorrem no início do mês, enquanto as de agosto no final. Esse padrão indica que climas mais frios têm uma relação negativa com a atividade de enxameação.

Outro fator que diferencia os dois períodos de capturas é a umidade média apresentada na Figura 3.8 (c), os meses com menores níveis de umidade média, são agosto e setembro, que

Figura 3.8 – Diagramas de rosa representando a época das capturas e variáveis ambientais ao longo dos meses: (a) Precipitação, (b) Temperatura Média, (c) Umidade Média, (d) Velocidade do Vento.



Fonte: Do autor (2025).

também são os períodos de maior número de capturas. Na Figura 3.8 (d), é possível observar que os meses de agosto e setembro registram as maiores velocidades de vento.

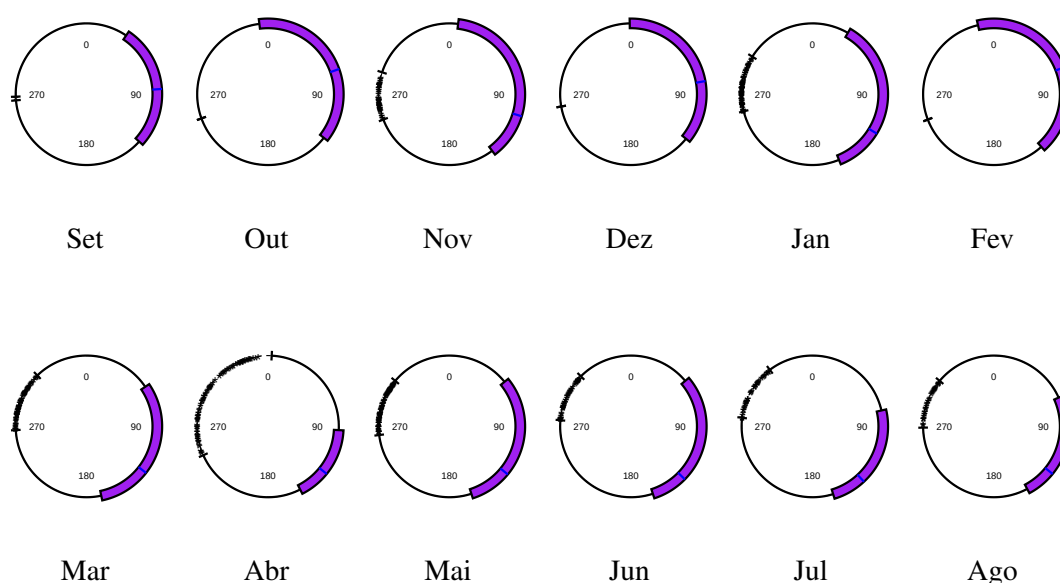
Em resumo, a presença de chuva está associada à ausência de capturas de enxames de abelhas africanizadas, e mesmo entre os meses secos, aqueles com temperaturas mais frias, como junho, julho e agosto, também apresentam ausência de capturas. Dentre os dois períodos de capturas identificados, o que abrange agosto e setembro destaca-se com o maior número de capturas, possivelmente potencializados pelos baixos valores de umidade média e as maiores velocidades de vento.

3.2.4.2 Direção do vento

A direção do vento, por sua natureza circular, foi analisada separadamente. Elaborou-se um boxplot circular mensal (Figura 3.9), que revela uma tendência predominante de ventos nos

primeiro e segundo quadrantes (Nordeste e Sudeste, respectivamente), independentemente do mês. Segundo Pereira *et al.* (2021), a direção e a intensidade dos ventos, juntamente com a localização geográfica, influenciam a ocorrência de precipitação, sendo a direção do vento uma das variáveis lineares mais determinantes para a captura de enxames.

Figura 3.9 – Boxplot circular da direção do vento (em graus) pelos meses do ano.



Fonte: Do autor (2025).

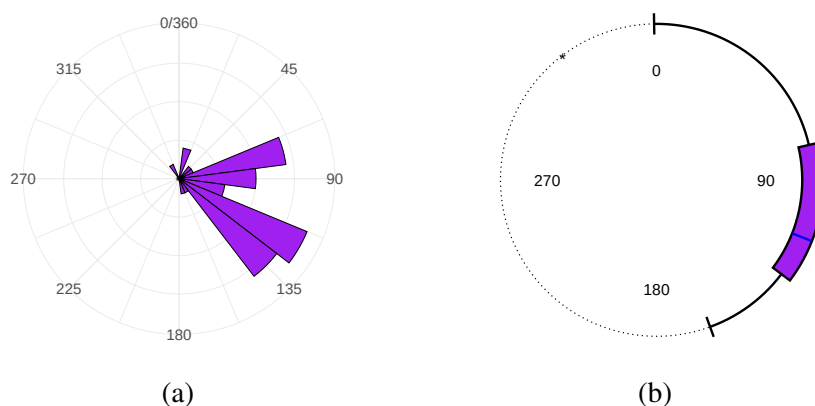
Para uma análise mais detalhada, foram construídos um diagrama de rosas, baseado nas médias diárias da direção do vento durante as capturas em ninhos-armadilhas (Figura 3.10 (a)), e um boxplot circular para os mesmos dados (Figura 3.10 (b)). Esses gráficos indicam uma preferência pela direção leste, sugerindo uma tendência consistente nas condições de captura.

3.2.5 Ângulo do alvado

A análise do ângulo do alvado, que representa a orientação da entrada dos ninhos-armadilha, foi realizada utilizando técnicas de estatística circular devido à sua natureza angular. Na Figura 3.11, são apresentados os limites do boxplot circular dos ângulos do alvado considerando os grupos analisados.

No grupo Instalados, observa-se uma distribuição por todos os ângulos, exibindo que os ninhos-armadilha foram distribuídos de maneira uniforme. Nos ninhos em que não houve captura (Ausente) também apresentam uniformidade em suas distribuições, com alguns outliers em ambos os casos completando a distribuição. No entanto, o grupo Captura Presente destaca-

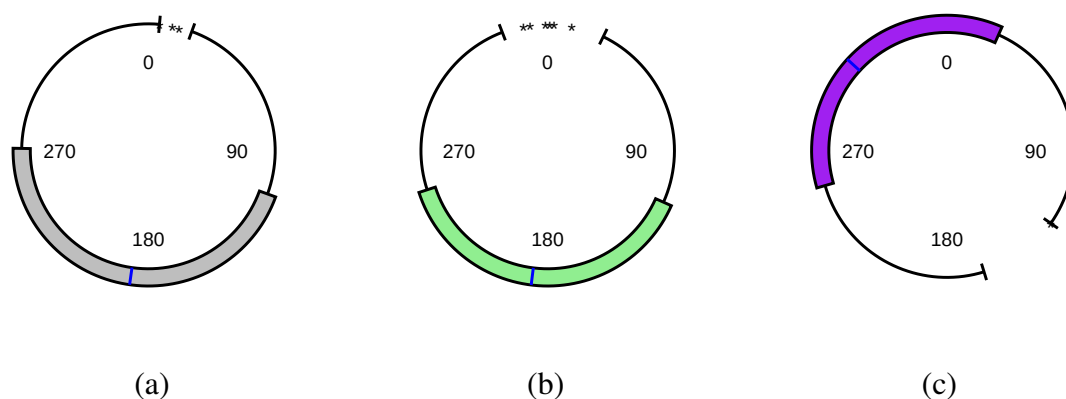
Figura 3.10 – Diagramas da direção do vento (em graus) durante os períodos de captura: (a) diagrama de rosas mostrando a distribuição circular das direções do alvado, (b) boxplot circular destacando a mediana, quartis e *outliers*.



Fonte: Do autor (2025).

se por exibir alteração no ângulo do alvado, com ausência no intervalo próximo ao intervalo de 90° a 180° no terceiro quadrante, indicando que o ângulo do alvado dos ninhos em que houve captura está localizados preferencialmente contrário a direção dos ventos conforme Figura 3.9.

Figura 3.11 – Boxplots circulares de alvado para os grupos (a) Instalados, (b) Ausente e (c) Presente, em graus.

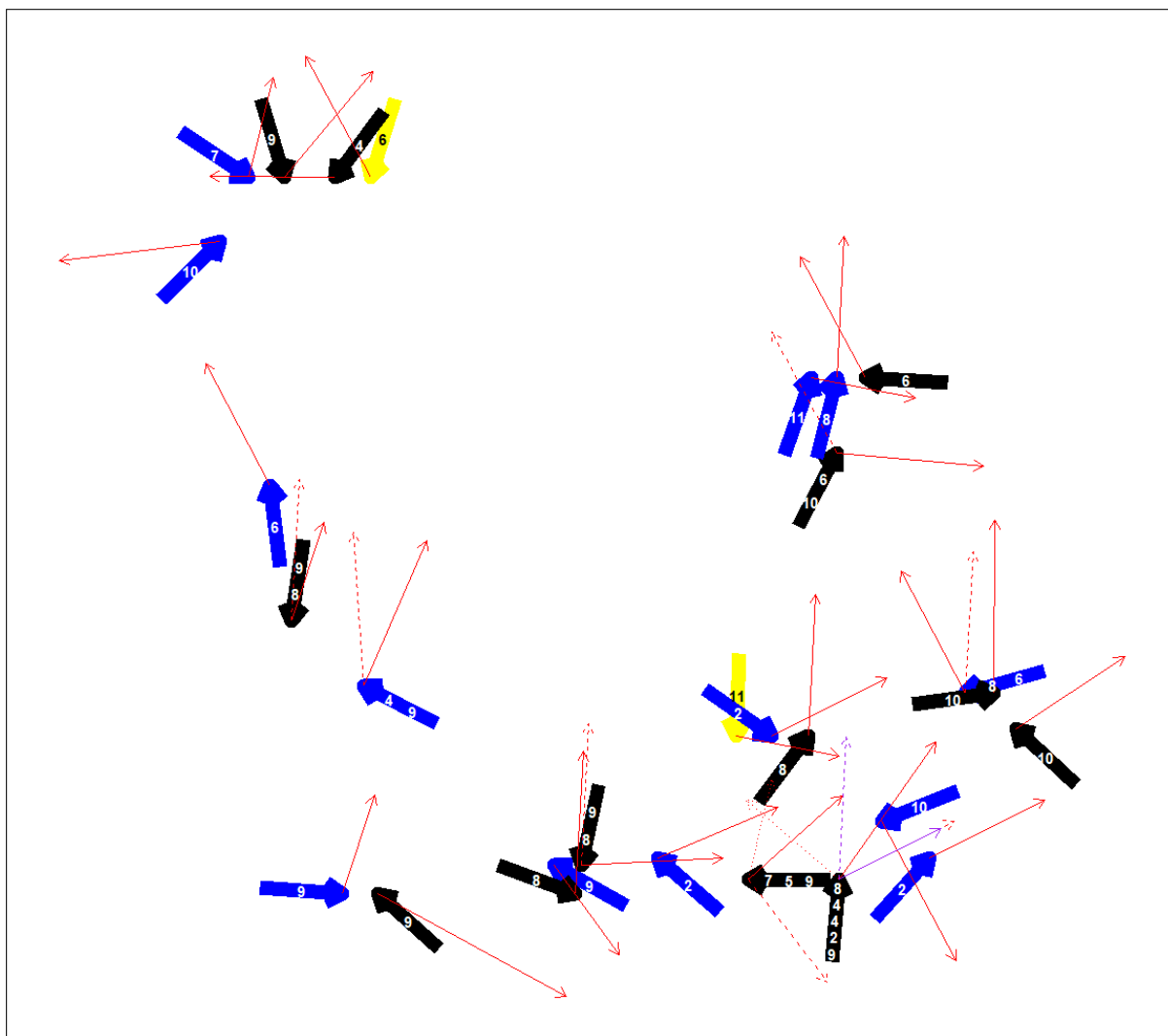


Fonte: Do autor (2025).

Pode-se dizer ainda que os boxplots circulares da Figura 3.11 revelam que a distância interquartílica dos grupos Instalados e Ausente é ampla, próxima a 180° , refletindo uma dispersão angular considerável e sem forte direcionalidade. No grupo Presente, a distância interquartílica é um pouco mais concentrada, embora ainda superior a 90° , indicando uma direção levemente preferencial em torno de 297° (noroeste).

A situação individual dos ninhos-armadilha em relação a cor de revestimento, indicação do mês de capturas/recapturas, direção e velocidade dos ventos pode ser observada na Figura

Figura 3.12 – Mapa dos ninhos-armadilhas com capturas e direção do vento, com pontos coloridos indicando as cores dos ninhos-armadilhas (azul, preto e amarelo) e flechas representando a direção e velocidade média do vento nos dias de captura.



Fonte: Do autor (2025).

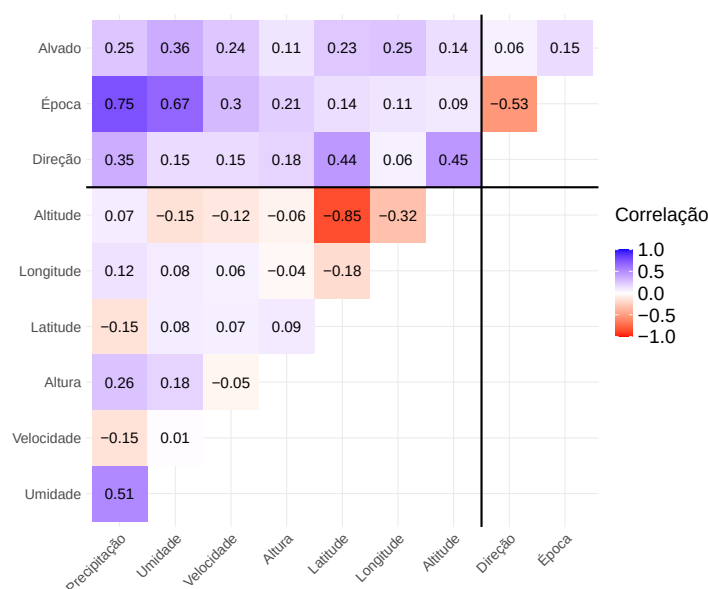
3.12. Que mostra que em sua maioria, o ângulo do vento é diferente do ângulo em que o alvado foi posicionado.

Para quantificar a relação existente entre as variáveis, realizou-se o estudo das correlações linear-linear, circular-circular e circular-linear.

3.3 Análise de Correlação entre Variáveis Ambientais e de Captura

Para a analisar as correlações entre as variáveis ambientais e as características associadas às capturas de enxames de abelhas africanizadas na área experimental, utilizou-se combinações

Figura 3.13 – Mapa de calor da matriz de correlação entre variáveis ambientais e de captura de enxames, segmentado em quatro regiões: linear-linear, circular-circular, linear-circular e circular-linear.



Fonte: Do autor (2025).

de correlações lineares e circulares para explorar as interações entre essas variáveis, organizando os resultados em um mapa de calor, ilustrado na Figura 3.13, que separa as relações em três categorias: (a) linear-linear, (b) circular-circular e (c) circular-linear.

A maior associação identificada é a correlação linear-circular entre precipitação total e época de captura (0,75), evidenciando que a ocorrência de chuvas está fortemente ligada à estação do ano, refletindo um padrão sazonal claro. A segunda correlação positiva mais relevante é a correlação linear-circular entre umidade média e época de captura (0,67), que destaca a relação entre níveis de umidade e sazonalidade, sugerindo que a umidade varia de forma consistente ao longo do ano, influenciada diretamente pela dinâmica climática anual.

A correlação linear-linear entre precipitação total e umidade média (0,51) indica uma associação positiva, em que maiores acumulados de chuva coincidem com níveis mais elevados de umidade, um resultado esperado devido à contribuição da precipitação para a umidade atmosférica, especialmente em escalas temporais curtas e em condições de alta pluviosidade. A correlação circular-circular entre direção do vento e época de captura (-0,53) indica uma associação, sugerindo que as direções predominantes do vento variam sazonalmente. Essas in-

terações destacam a complexidade das relações entre fatores climáticos e sazonais, reforçando a importância de considerar tanto variáveis lineares quanto circulares para otimizar a instalação de ninhos-armadilhas.

Já a correlação linear-circular entre altitude e direção do vento (0,45) aponta que mudanças na direção do vento podem influenciar enxames de Abelhas africanizadas a procurar ninhos em diferentes altitudes. A correlação linear-linear entre altitude e latitude (-0,85) revela uma forte relação negativa, explicada pelo contexto geográfico do experimento, realizado na Fazenda Experimental Risoleta Neves, localizada em uma região montanhosa no *Campus* Tancredo de Almeida Neves da UFSJ. Como o estudo foi conduzido em declive, a latitude diminui à medida que a altitude aumenta, refletindo a inclinação do terreno. A correlação linear-circular entre direção do vento e latitude (0,44) apresenta uma associação relativamente alta, que pode ser explicada pelo contexto geográfico e topográfico do experimento realizado na Fazenda Experimental Risoleta Neves, em uma região montanhosa. Como previamente discutido, a latitude está fortemente correlacionada negativamente com a altitude (-0,85) devido à disposição dos ninhos-armadilhas ao longo da encosta da montanha, onde latitudes menores correspondem a altitudes maiores, mais ao norte, e latitudes maiores a altitudes menores, mais ao sul.

4 CONCLUSÕES

O sucesso da captura de enxames de abelhas africanizadas foi avaliado por meio da análise descritiva das variáveis lineares e circulares identificando-se um padrão bimodal que influencia a atratividade de ninhos-armadilhas.

A integração de técnicas estatísticas descritivas e de correlação, adaptadas à natureza linear ou circular das variáveis, permitiu quantificar a influência entre fatores ambientais e de posicionamento dos ninhos.

Os resultados evidenciaram que a cor do revestimento dos ninhos-armadilhas desempenha um papel relevante na atratividade, com ninhos pretos e azuis apresentando maior sucesso em capturas em comparação com os amarelos. Essa preferência pode estar associada à percepção visual das abelhas ou a condições ambientais.

A altura de instalação também foi mais elevada nos ninhos em que houve captura, possivelmente devido à maior exposição a ventos.

A análise sazonal revelou dois períodos principais de enxameação, concentrados no outono (10 de abril a 06 de junho) e na primavera (29 de agosto a 19 de outubro), caracterizando a bimodalidade das capturas. Esses períodos coincidem com condições climáticas favoráveis, marcadas por baixa precipitação, temperaturas mais elevadas e menor umidade, especialmente em agosto e setembro, que registraram o maior número de capturas. A ausência de capturas em meses chuvosos, como dezembro e janeiro, reforça a influência negativa da precipitação na atividade de enxameação.

As variáveis circulares, como a orientação do alvado e a direção do vento, apresentaram padrões distintos. A direção do vento concentrou-se predominantemente nos quadrantes nordeste e sudeste, enquanto os ângulos do alvado nos ninhos com capturas tiveram uma preferência por orientações opostas à direção predominante dos ventos. Essa relação pode estar ligada à proteção contra correntes de ar ou à facilitação da entrada e saída das abelhas.

A análise de correlação destacou algumas interações, como a forte associação entre precipitação e época de captura (0,75) e entre umidade e sazonalidade (0,67), refletindo a influência do clima na dinâmica de enxameação. Portanto, este estudo demonstra que o sucesso na captura de enxames de abelhas africanizadas depende de uma combinação de fatores ambientais e de configuração dos ninhos-armadilhas. Recomenda-se que apicultores priorizem ninhos pretos ou azuis, instalados a alturas superiores a 1,70 m, com orientação do alvado contrária à direção pre-

dominante do vento, e posicionados em períodos de baixa precipitação, como o final do inverno e início da primavera. Essas estratégias podem otimizar a captura de enxames, contribuindo para a sustentabilidade da apicultura em regiões com características climáticas e topográficas semelhantes às do presente estudo.

REFERÊNCIAS

- ASUERO, A. G.; SAYAGO, A.; GONZÁLEZ, A. The correlation coefficient: An overview. **Critical reviews in analytical chemistry**, Taylor & Francis, v. 36, n. 1, p. 41–59, 2006.
- BARUQUI, A. M. a. *et al.* Levantamento de reconhecimento de média intensidade dos solos da zona campos das vertentes-mg. 2006.
- BEIER, W.; LINDAUER, M. Der sonnenstand als zeitgeber für die biene. **Apidologie**, v. 1, n. 1, p. 5–28, 1970.
- CAPALDI, E. A.; DYER, F. C. The role of orientation flights on homing performance in honeybees. **Journal of Experimental Biology**, v. 202, n. 12, p. 1655–1666, 1999.
- COUVILLON, M. J.; SCHÜRCH, R.; RATNIEKS, F. L. Waggle dance distances as integrative indicators of seasonal foraging challenges. **PloS one**, Public Library of Science San Francisco, USA, v. 9, n. 4, p. e93495, 2014.
- ESQUIVEL, D. M. S. *et al.* Do geomagnetic storms change the behaviour of the stingless bee guirucu (schwarziana quadripunctata)? **Naturwissenschaften**, v. 94, p. 139–142, 2007.
- GOOGLE. **Google Maps**. 2025. Disponível em: <https://www.google.com/maps>. Acesso em: 9 jun. 2025.
- GOULD, J. L.; KIRSCHVINK, J. L.; DEFFEYES, K. S. Bees have magnetic remanence. **Science**, v. 201, n. 4360, p. 1026–1028, 1978.
- JAMMALAMADAKA, S.; SENGUPTA, A. Circular statistics with applications. **Journal of Statistical Planning and Inference**, Elsevier, v. 99, n. 1, p. 1–20, 2001.
- LACERDA, D. C. O. *et al.* Influência dos pontos cardeais e colaterais na nidificação de abelhas nativas em colmeias octogonais. **Gaia Scientia**, v. 11, n. 2, p. 203–217, 2017.
- LEVEN, L. *et al.* A apicultura nas regiões tropicais. **Série Agrodok No. 32**, Agromisa, 2006.
- LUCANO, M. J. **Caracterização de material magnético mineralizado em abelhas meliponini Schwarziana quadripunctata: magnetometria SQUID e Ressonância Ferromagnética**. 33–37 p. Tese (Doutorado) — Centro Brasileiro de Pesquisas Físicas, 2006.
- MACEDO, C. R. C. *et al.* Comportamento da nidificação de abelhas melíponas. **Ciência Animal Brasileira**, v. 21, 2020.
- MARDIA, K. Linear-circular correlation coefficients and rhythmometry. **Biometrika**, JSTOR, p. 403–405, 1976.
- NETTO, M. M. G. *et al.* Entre a serra e o rio: uma abordagem territorial. **A Morfologia Urbana de Tiradentes/MG**, p. 22, 2023.
- PEREIRA, F. d. M. *et al.* **Criação de abelhas (Apicultura)**. 2016.
- PEREIRA, G. *et al.* Análise da variabilidade da precipitação para o estado de minas gerais (1981-2017). **Revista Brasileira de Climatologia**, v. 1, fev. 2021. Disponível em: <https://ojs.ufgd.edu.br/rbclima/article/view/14068>.

PINTO, F. de A.; SILVA, E. C. da; ANJOS, A. J. dos. Influências climáticas no perfil sanitário em colônias de apis mellifera africanizadas. **Revista Técnica Ciências Ambientais**, v. 1, n. 8, 2024.

QU, Y. *et al.* Biology of bamboo nesting stingless bees of southwest china. **Apidologie**, Springer, v. 53, n. 5, p. 53, 2022.

SANTOS, W. R. D.; CARNEIRO, T. C. J.; OLIVEIRA, M. P. V. D. Consumidor de dispositivos eletrônicos portáteis: O efeito moderador do perfil do adotante na intenção de compra. **ReMark-Revista Brasileira de Marketing**, v. 15, n. 4, p. 496–511, 2016.

SILVA, I. S. *et al.* Comparação da produção de abelhas da espécie apis mellifera com as mudanças ambientais. **Animal em Foco**, v. 3, p. 68–76, 2023.

SOUSA, D. S. R.; RIZZARDO, R. A. G. Dinâmica populacional de colônias de apis mellifera durante o período chuvoso na região de araguaína. **Revista Brasileira de Zootecnia**, Universidade Federal do Tocantins, Araguaína, TO, Brasil, v. 43, n. 5, p. 250–256, 2014.

STEFFAN-DEWENTER, I.; KUHN, A. Honeybee foraging in differentially structured landscapes. **Proceedings of the Royal Society of London. Series B: Biological Sciences**, The Royal Society, v. 270, n. 1515, p. 569–575, 2003.

Topographic Map. **Mapa topográfico do campus CTAN da UFSJ - São João del-Rei**. 2025. <https://pt-br.topographic-map.com/map-mvwgt/S%C3%A3o-Jo%C3%A3o-del-Rei/?zoom=17&cer=-21.10462,-44.25132>. Acessado em 04 de abril de 2025.

VAZ, M. A. *et al.* Comportamento de nidificação de melipona subnitida (ducke, 1910) e frieseomelitta sp. no seridó oriental do rio grande do norte, brasil. **Research, Society and Development**, v. 10, n. 8, 2021.

VINÍCIUS-SILVA, R.; FERRAZ, N. P.; WERNECK, M. do V. Levantamento da fauna de abelhas (hymenoptera, apoidea) em um fragmento de mata atlântica, no município de ubá, minas gerais: Survey of the bee fauna (hymenoptera, apoidea) in an atlantic forest fragment at ubá municipality, minas gerais state, southeastern brazil. **Latin American Journal of Development**, v. 3, n. 5, p. 3203–3216, 2021.

Artigo 4 - Elaborado conforme a norma NBR 6022	
Título do artigo:	Análise de sobrevivência da captura de enxames de abelhas africanizadas com eventos recorrentes
Autores:	Luiz Fernando Silva Resende Carla Regina Guimarães Brighenti



**ARTIGO 4 – ANÁLISE SOBREVIVÊNCIA NA CAPTURA DE ENXAMES DE
ABELHAS AFRICANIZADAS COM EVENTOS RECORRENTES
SURVIVAL ANALYSIS IN THE CAPTURE OF AFRICANIZED BEE SWARMS
WITH RECURRENT EVENTS¹**

Luiz Fernando Silva Resende

<https://orcid.org/0000-0003-4697-6503>

Carla Regina Guimarães Brighenti

<https://orcid.org/0000-0002-7822-3744>

Resumo: O modelo de Andersen-Gill para eventos recorrentes foi aplicado para examinar o tempo até a captura de enxames de abelhas africanizadas em ninhos-armadilha na Fazenda Experimental Risoleta Neves, São João del Rei, de setembro de 2023 a agosto de 2024. Avaliaram-se variáveis categóricas (cor), lineares (altura, altitude) e circulares (alvado, localização, época). Constatou-se bimodalidade nas épocas de capturas, com picos no pós-chuva (abril-junho) e pré-chuva (agosto-setembro), associados a baixa precipitação e temperaturas elevadas. Concluiu-se que, no pós-chuva, ninhos pretos e orientação do alvado otimizam capturas, pois as abelhas, em temperaturas mais baixas, buscam posicionar as caixas para aumentar a captação solar e elevar a temperatura; no pré-chuva, cores escuras (preta e azul), orientando práticas apícolas sustentáveis.

Palavras-chave: Andersen-Gill; bimodalidade; práticas apícolas; estatística circular; ninhos-armadilha.

Abstract: The Andersen-Gill model for recurrent events was applied to evaluate the time to capture Africanized bee swarms in trap nests at Risoleta Neves Experimental Farm, São João del Rei, from September 2023 to August 2024. Categorical (color), linear (height, altitude), and circular (entrance angle, location, season) variables were assessed. A bimodal pattern was found in capture periods, with peaks in post-rain (April-June) and pre-rain (August-September), linked to low precipitation and higher temperatures. It was concluded that, in the post-rain period, black nests and entrance angle orientation optimize captures, as bees in cooler temperatures position nests to enhance solar capture and increase temperature; in the pre-rain period, dark colors (black and blue), guiding sustainable apicultural practices.

Keywords: Andersen-Gill model; bimodality; apicultural practices; circular statistics; trap nests.

¹ Artigo científico apresentado à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária.

1 INTRODUÇÃO

A análise de sobrevivência é uma área da estatística que se dedica ao estudo do tempo até a ocorrência de um evento de interesse, frequentemente lidando com dados censurados, nos quais a informação sobre o evento pode ser parcial ou incompleta (Turkson; Ayiah-Mensah; Nimoh, 2021). Essa abordagem é amplamente aplicada em diversas áreas, como medicina, engenharia, trânsito e ecologia, devido à sua capacidade de modelar fenômenos complexos nos quais o tempo de espera é influenciado por múltiplos fatores (Liang *et al.*, 2022; RASHEED, 2023; Tesfay *et al.*, 2025). Na apicultura, a análise de sobrevivência oferece uma técnica adequada para investigar padrões temporais associados à captura de enxames, um processo que apresenta desafios únicos devido à interação entre fatores biológicos, ambientais e geográficos (Brodschneider; Gray, 2022).

A captura de enxames de abelhas, especialmente das africanizadas (*Apis mellifera*), é uma prática essencial na apicultura, pois permite a multiplicação de colmeias de forma sustentável e de baixo custo (Sultana *et al.*, 2024). Esse processo ocorre predominantemente durante períodos de enxameação, quando colônias se dividem em busca de novos locais para nidificação, motivadas por condições como escassez de recursos florais, superpopulação ou mudanças climáticas (Mouillard-Lample *et al.*, 2023). Fatores como a disponibilidade de floradas, temperatura, precipitação e características dos ninhos-armadilha, incluindo altura, cor e orientação, influenciam diretamente a atratividade desses locais para os enxames (Césard, 2022; Renteria; Brehm, 2025). Compreender os critérios que determinam a escolha de um ninho-armadilha por um enxame é de grande interesse para apicultores, pois otimiza o manejo e aumenta a eficiência da captura (Antoine; Forrest, 2021).

Acompanhar a captura de enxames de abelhas africanizadas não é um procedimento trivial, pois envolve uma complexa interação de fatores que influenciam o sucesso das instalações dos ninhos-armadilhas. Esses dispositivos, geralmente construídos com materiais acessíveis como caixas revestidas e posicionados estrategicamente em árvores ou outras estruturas, devem atender a condições específicas para atrair os enxames (Rahimi; Barghjelveh; Dong, 2021). Embora a literatura aborde aspectos como altura, localização e condições ambientais, não há consenso sobre os principais fatores que determinam a escolha do posicionamento pelos enxames. Estudos comparativos, como os realizados com ninhos de pássaros, sugerem que a orientação em relação ao sol pode desempenhar um papel significativo na atratividade (Lam-

brechts; Deeming, 2024). Além disso, a cor do revestimento dos ninhos-armadilhas é outro fator de interesse, a exemplo do uso de cores claras, como branco e amarelo, em armadilhas para formigas (Akkuzu *et al.*, 2021).

O posicionamento dos ninhos-armadilhas pode ser influenciado pela disponibilidade de luz e pela posição do sol, que, por sua vez, afetam diretamente o ângulo de entrada do ninho, conhecido como alvado. Esse ângulo está intrinsecamente relacionado às coordenadas geográficas de latitude e longitude, bem como à orientação magnética, que pode orientar as abelhas na escolha de locais para nidificação. Assim, a determinação precisa do ângulo do alvado é essencial para compreender os padrões de captura, exigindo o uso de técnicas de estatística circular para modelar adequadamente a natureza cíclica dessas variáveis (Vale; Acosta-Avalos, 2021). A análise circular permite capturar a periodicidade inerente aos dados angulares, fornecendo uma representação robusta das preferências de orientação dos enxames em relação a fatores ambientais e geográficos (Siehler; Wang; Bloch, 2021).

Essas capturas podem envolver múltiplas ocorrências, já que após uma captura, o ninho-armadilha com enxame pode ser substituído por um novo ninho vazio, com as mesmas características, recolocados na mesma posição, caracterizando eventos recorrentes que demandam modelos estatísticos capazes de capturar a dependência temporal entre essas capturas. Nesse contexto, Farooq *et al.* (2025) utilizaram o modelo Andersen-Gill em uma análise de sobrevivência para estudar o tempo entre o estabelecimento do mosquito *Aedes albopictus* e surtos recorrentes de dengue e chikungunya em regiões da União Europeia. O estudo revelou que características climáticas e geográficas, aceleram a recorrência de surtos.

Na apicultura, a reposição de ninhos-armadilha pode resultar em capturas recorrentes, influenciadas por variáveis como orientação do alvado, cor de revestimento, localização geográfica e sazonalidade, que apresentam natureza cíclica e requerem a integração de estatística circular para modelar suas propriedades angulares. Diante do exposto, o objetivo deste trabalho é integrar a estatística circular à análise de sobrevivência, utilizando o modelo de Andersen-Gill para avaliar a recaptura bem como as covariáveis que influenciam o tempo até a captura de enxames de abelhas africanizadas.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta os fundamentos teóricos que embasam a análise de sobrevivência aplicada ao estudo do tempo de captura de enxames de abelhas africanizadas, com ênfase em eventos recorrentes e modelos de regressão.

2.1 Dados de Sobrevivência

A análise de sobrevivência estuda o tempo de sobrevivência T até a ocorrência de um evento de interesse, frequentemente lidando com dados incompletos devido à censura (Turkson; Ayiah-Mensah; Nimoh, 2021). A variável aleatória T é contínua e não negativa, acompanhada por uma variável indicadora $Y_i(t)$, que determina se o indivíduo i está sob risco no instante t :

$$Y_i(t) = \begin{cases} 0, & \text{se o indivíduo } i \text{ não está sob risco no instante } t; \\ 1, & \text{se o indivíduo } i \text{ está sob risco no instante } t. \end{cases}$$

Essa variável indicadora é registrada como uma informação adicional, assim como possíveis covariáveis. Pode-se definir três funções para descrever os diferentes aspectos apresentados pelos dados de sobrevivência (Tomazella, 2003):

- a) função densidade de probabilidade: a função densidade $f(t) > 0$ para todo t maior que zero é definida como o limite da probabilidade da ocorrência do evento de um indivíduo no intervalo de tempo $[t, t + dt)$, expressa por (Cox; Oakes, 1984):

$$f(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T \leq t + dt)}{dt};$$

- b) função de sobrevivência: a função $S(t)$ descreve a forma distribucional dos tempos de sobrevivência por meio da probabilidade da não ocorrência do evento de interesse até o instante t , sendo denominada função de sobrevivência. É uma função não crescente, com $S(0) = 1$ e $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$, dada por (Lawless, 1982):

$$S(t) = P(T \geq t);$$

- c) função de risco: a função de risco, ou taxa de falha, é definida pelo limite da probabilidade de um indivíduo falhar no intervalo $[t, t + dt)$, dado que não tenha falhado até o instante t , expressa por (Cox; Oakes, 1984):

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt \mid T \geq t)}{dt} = \frac{f(t)}{S(t)}.$$

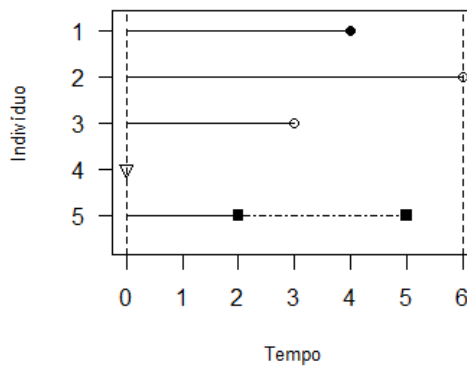
2.2 Mecanismo de Censura

A censura ocorre quando o tempo exato do evento não é observado, resultando em informação parcial (Andersen *et al.*, 2021). A Figura 2.1 representa os principais tipos de censura, que são:

- a) censura à direita: o evento não é observado até o final do estudo, seja por término do período de observação, número fixo de eventos ou retirada por causas externas;
- b) censura à esquerda: o evento ocorreu antes do início do estudo, mas o momento exato é desconhecido;
- c) censura intervalar: o evento é conhecido por ter ocorrido em um intervalo $[T_0, T_F]$, no qual T_0 é o último momento sem evento e T_F é o momento em que o evento é verificado.

Figura 2.1 – Representação gráfica de registros de tempos com censura e sem censura.

Legenda: (- -) pontilhado vertical representa o início e fim do estudo; (—) retas horizontais corresponde ao período que o indivíduo esteve sob risco; (●) a ocorrência do evento; (○) censura à direita; (▽) censura à esquerda; (- · -) pontilhado horizontal representa a censura intervalar com (■) sendo os intervalos de início e fim.



Fonte: Do autor (2025).

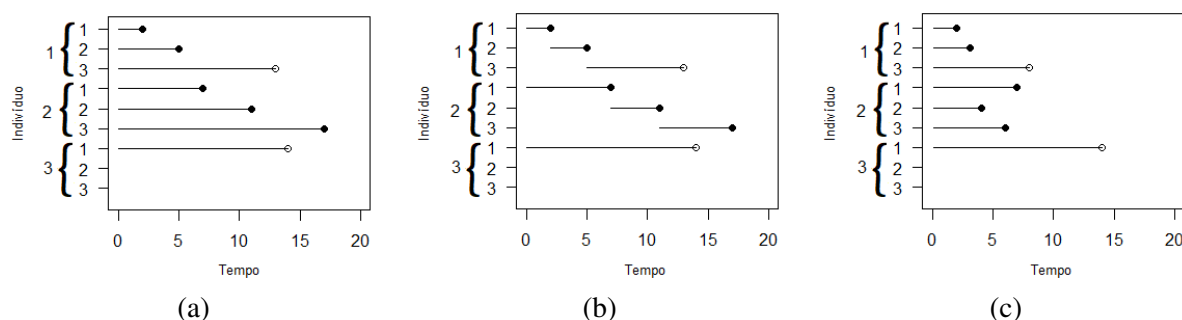
Na Figura 2.1, somente o indivíduo 1 teve a ocorrência do evento antes do fim do estudo, no tempo $t = 4$. O indivíduo 2 não teve ocorrência do evento durante o prazo estabelecido pelo estudo. Caso o prazo tenha sido determinado por uma unidade de tempo pré-estabelecida, trata-se de censura à direita do tipo I; se o prazo tiver sido determinado por uma quantidade pré-determinada de falhas, então se trata de censura à direita do tipo II. O indivíduo 3 foi retirado do experimento por um motivo diferente do evento de interesse, configurando uma censura à direita aleatória. O indivíduo 4 já havia experimentado o evento ao iniciar o estudo, representado pela censura à esquerda no tempo $t = 0$. O indivíduo 5 não teve a ocorrência de seu evento marcada no exato momento da falha, mas sabe-se que ocorreu entre os tempos $T_0 = 2$ e $T_F = 5$. Também é possível afirmar que o indivíduo 5 não foi avaliado nos tempos $t = 3$ e $t = 4$, uma vez que o último tempo em que se sabe que não houve ocorrência da falha foi em $T_0 = 2$, e o instante em que foi verificado que a falha já havia ocorrido foi em $T_F = 5$.

2.3 Eventos Recorrentes

Eventos recorrentes referem-se à ocorrência repetida do mesmo evento em uma unidade experimental, como múltiplas capturas de enxames em um mesmo ninho-armadilha (Chiou *et al.*, 2023). Esses eventos requerem modelos que considerem a estrutura temporal das ocorrências, permitindo acomodar tanto a independência quanto a dependência entre os eventos, conforme capturado pelos registros de tempo t_{ij} para o i -ésimo objeto e j -ésima ocorrência. Os intervalos sob risco podem ser registrados por diferentes métodos, como ilustrado na Figura 4.1, que apresenta representações gráficas dos métodos Total Time (TT), Counting Process (CP) e Gap Time (GT). Seguindo (Kelly; Lim, 2000), denotamos o tempo t_{ij} como o tempo da j -ésima ocorrência do evento para o i -ésimo objeto de estudo, onde i indexa os ninhos-armadilhas e j indexa os eventos recorrentes.

- a) total Time (TT): assume independência entre as ocorrências j , com intervalo de risco $[0, t_{ij}]$, onde t_{ij} é o tempo da j -ésima ocorrência em i ;
- b) counting Process (CP): considera dependência entre os eventos, assumindo que o tempo inicial $t_{i,j+1}$ ocorre somente após o tempo final $t_{i,j}$;
- c) gap Time (GT): reinicia o tempo após cada ocorrência, com intervalo de risco $[0, g_{ij}]$, onde g_{ij} é o intervalo entre ocorrências consecutivas do sujeito i , dado por $g_{ij} = t_{ij} - t_{i,j-1}$, modelando a dependência temporal.

Figura 2.2 – Representação gráfica de registros dos tempos: (a) Total Time, (b) Counting Process e (c) Gap Time.



Fonte: Adaptado de Kelly; Lim (2000).

A Figura 2.2 (a), ilustra-se o método Total Time (TT), em que o primeiro indivíduo apresenta o primeiro evento no intervalo $(0,2)$, o segundo no intervalo $(0,5)$ e uma censura em $(0,13)$. O segundo indivíduo registra o primeiro evento em $(0,7)$, o segundo em $(0,11)$ e o terceiro em $(0,17)$. O terceiro indivíduo apresenta apenas uma censura no intervalo $(0,14)$.

No método Counting Process (CP) (Figura 2.2 (b)) o final dos intervalos é idêntico ao método TT, porém o intervalo sob risco de um evento é registrado a partir da ocorrência do evento anterior. Já no método Gap Time (GT) (Figura 2.2 (c)) os intervalos tem o mesmo comprimento do método GT, porém todos iniciando em 0.

2.4 Modelos na Análise de Sobrevida

Modelos de regressão na análise de sobrevivência são ferramentas fundamentais para avaliar como covariáveis influenciam o tempo até a ocorrência de um evento de interesse, sendo divididos em três categorias principais: não paramétricos, paramétricos e semiparamétricos (Selingerova; Katina; Horova, 2021). Esses modelos permitem capturar relações entre variáveis explicativas e a probabilidade de sobrevivência, adaptando-se a diferentes contextos e características dos dados.

Os modelos não paramétricos oferecem uma abordagem sem suposições sobre a distribuição do tempo de sobrevivência. Um exemplo clássico é o estimador de Kaplan-Meier, que calcula a função de sobrevivência $S(t)$ diretamente a partir dos tempos de evento e censura observados (Kaplan; Meier, 1958). Esse não consegue estimar diretamente funções de sobrevivência usando dados censurados por intervalo. Para resolver isso, o método de Turnbull (Turnbull, 1976) foi desenvolvido como uma abordagem não paramétrica para estimar funções

de sobrevivência diretamente a partir de dados censurados por intervalo (Iijima; Momozaki; Ando, 2025).

O modelos paramétricos são aqueles em que $S(t)$ é determinada por parâmetros com distribuições de probabilidades conhecidas. As vantagens de se presumir que os dados são provenientes de uma determinada distribuição de probabilidade são as inferências que podem ser feitas sobre os parâmetros da distribuição (Moncharmon *et al.*, 2003). Para estimar os parâmetros, na presença de censura, há um consenso na utilização do método de máxima verossimilhança (Silva, 2016),

2.5 Modelo Semiparamétrico de Cox

O modelo de Cox (Cox, 1972), um modelo semiparamétrico, assume que o risco $\lambda(t)$ é proporcional entre indivíduos, expresso como:

$$\lambda(t_i | x_i) = \lambda_0(t_i) \exp(x_i' \beta),$$

em que $\lambda_0(t)$ é o risco basal, uma função não paramétrica e não negativa, x_i representa as covariáveis do sujeito i , e β é o coeficiente das covariáveis, parte paramétrica. A função de verossimilhança parcial é usada para estimar β , sem necessidade de modelar $\lambda_0(t)$ (Chung *et al.*, 2018).

2.6 Extensão do Modelo de Cox para dados recorrentes: Modelo de Andersen-Gill

A diferença entre o modelo de Cox clássico, descrito anteriormente, e os modelos para eventos recorrentes está na definição do conjunto de indivíduos em risco. No modelo clássico, o indivíduo deixa de estar em risco após experimentar o evento. Para eventos recorrentes, porém, o indivíduo permanece em risco após cada ocorrência do evento de interesse (Tomazella, 2003).

Diversas extensões de modelos de sobrevivência baseadas na abordagem de riscos proporcionais de Cox foram propostas para lidar com dados de eventos recorrentes (Kelly; Lim, 2000), sendo uma delas o modelo de Andersen-Gill (AG) (Andersen; Gill, 1982). O modelo AG considera que todas as ocorrências de um sujeito contribuem para o conjunto de risco, independentemente do número de eventos anteriores, utilizando uma estrutura não estratificada.

A função de risco é expressa como:

$$\lambda(t_{ij} | x_i) = \lambda_0(t_{ij}) \exp(x_i' \beta),$$

onde t_{ij} é o tempo da j -ésima ocorrência para o i -ésimo sujeito, conforme definido anteriormente, x_i são as covariáveis do sujeito i , e $\lambda_0(t_{ij})$ é o risco basal comum, não estratificado, ou seja, o mesmo para todas as ocorrências e sujeitos em cada tempo t_{ij} (Andersen; Gill, 1982; Kelly; Lim, 2000). Essa abordagem agrupa todas as recorrências de cada sujeito no conjunto de risco, permitindo a modelagem da dependência temporal entre eventos por meio de covariáveis e ajustes robustos de variância (Carvalho *et al.*, 2011).

A escolha do modelo AG, cujo o registro de tempo é o *counting process*, para este estudo justifica-se por sua adequação à análise de capturas recorrentes de enxames em ninhos-armadilhas, em que o ninho $j + 1$ só é colocado após a captura de enxame no ninho j .

3 METODOLOGIA

A metodologia utilizada para avaliar os fatores que influenciam o tempo até a captura de enxames de abelhas africanizadas em ninhos-armadilhas utiliza a análise de sobrevivência com foco no modelo de Andersen-Gill (AG) para eventos recorrentes. A abordagem combina variáveis lineares e circulares, considerando características dos ninhos e fatores sazonais. A metodologia abrange a definição das covariáveis, a estruturação dos dados de sobrevivência e a justificativa para a escolha do modelo AG em comparação com outras abordagens.

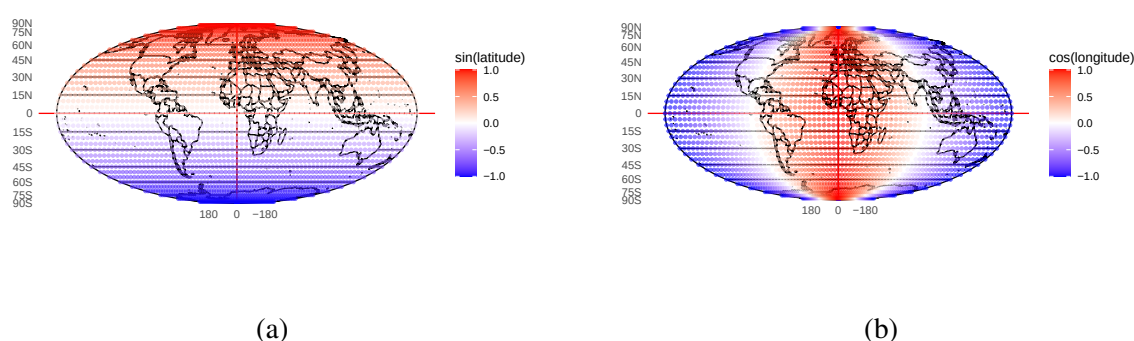
3.1 Caracterização do experimento e das Covariáveis

O experimento foi conduzido na Fazenda Experimental Risoleta Neves, *Campus* Tancredo de Almeida Neves (CTAN), Universidade Federal de São João del-Rei (UFSJ). Foram distribuídas 93 ninhos-armadilha em grupos de 3 em cada localidade, em uma área de 0,162 km^2 , sendo considerado para cada ninho as seguintes covariáveis:

- a) identificador único (*id*): cada ninho-armadilha foi associado a um identificador único (*id*), permitindo o rastreamento individual de suas características e garantindo a distinção entre eventos recorrentes no mesmo local ao longo do período experimental;
- b) cor do revestimento (*Cor*): variável categórica com três níveis: amarelo, azul e preto, aplicada em 31 ninhos-armadilhas por cor, totalizando as 93 unidades. A escolha das cores testa hipóteses sobre atratividade visual no modelo, utilizando a cor amarelo como referência, para estimar o efeito relativo de azul e preto no risco de captura;
- c) altitude (*Altitude*): variável contínua (metros acima do nível do mar), considerada por sua influência em microclimas locais, foi tratada como contínua no modelo;
- d) altura de instalação (*Altura*): variável contínua em metros, mantendo-se a escala original no modelo para capturar efeitos de posicionamento em relação ao solo;
- e) orientação do alvado (*Alvado*): variável circular em graus, 0° a 360°, indicando a direção da entrada do ninho em relação ao norte magnético, modelada usando $\sin(\text{Alvado})$ e $\cos(\text{Alvado})$ para incorporar sua natureza cíclica no modelo de Cox;

- f) localização geográfica (Latitude, Longitude): variáveis circulares, em graus, registradas conforme a natureza cíclica em $\sin(\text{Latitude})$ e $\cos(\text{Longitude})$ variando conforme a Figura 3.1. Para evitar instabilidade numérica e melhorar a interpretabilidade dos efeitos geográficos, elas foram normalizadas pela média devido à pequena área experimental (0,162 km²);
- g) época do ano (Época): variável circular (graus), representando o dia do ano (1 a 365) em escala angular (360°/365 por dia). Incluída para registrar a data de ocorrência das capturas e refletir a sazonalidade como períodos de enxameação influenciados por condições climáticas (Leven *et al.*, 2006). Foi modelada como $\sin(\text{Época})$ e $\cos(\text{Época})$ para refletir sua periodicidade.

Figura 3.1 – Visualização do seno da latitude (a) e cosseno da longitude (b) em uma grade global, em que as cores indicam uma variação de -1 (Vermelho) a +1 (Azul)



Fonte: Do autor (2025)

3.2 Entrada de Dados para Análise de Sobrevivência

Os dados de sobrevivência foram estruturados para modelar o tempo até a captura de enxames, considerando a possibilidade de eventos recorrentes (múltiplas capturas no mesmo local) e censura intervalar, devido à natureza periódica das observações. O processo envolveu:

- a) evento de interesse: a captura de um enxame de abelhas africanizadas em um ninho-armadilha, registrado como $\text{Captura} = 1$. A ausência de captura até o final do intervalo de observação ou devido a eventos adversos (e.g., ocupação por outras espécies, danos físicos) foi codificada como censura ($\text{Captura} = 0$);

- b) estrutura temporal: cada ninho-armadilha foi monitorado por 365 dias, realizando-se um total de 78 visitas, em cada uma das 93 caixas, no período entre 1 de setembro de 2023 a 31 de agosto de 2024. O tempo foi dividido em intervalos $[\text{Início}, \text{Fim}]$, no qual Início é o último dia sem captura e Fim é o dia da captura ou ocorrência de censura. Para recapturas, o ninho foi substituído, iniciando um novo intervalo (Estrato), codificado como 1 para o primeiro ninho, 2 para o substituto, e assim por diante.

As covariáveis consideradas foram:

- a) categóricas: Cor ;
- b) linear: Altitude e Altura ;
- c) circular: Alvado , Latitude , Longitude e Época .

3.3 Extensão do Modelo de Cox para dados recorrentes: Modelo de Andersen-Gill

O modelo de Cox, foi utilizado como base para avaliar o efeito das covariáveis no tempo até a captura de enxames, testando-se anteriormente o pressuposto de risco proporcionais através do teste de Shoenfeld.

O modelo de AG foi escolhido por sua adequação às características do estudo. O estudo permite múltiplas capturas no mesmo local (ninho-armadilha), com reposição após cada captura. O modelo AG acomoda essa estrutura, tratando cada evento como parte de um processo contínuo, ao contrário do modelo Cox padrão, que assume um único evento por unidade. Métodos como Total Time (TT) ignoram dependência entre eventos, enquanto Gap Time (GT) reinicia o tempo após cada evento, o que não reflete a continuidade do risco na área experimental.

O AG modela o risco basal como comum a todos os eventos, permitindo que covariáveis fixas influenciem o risco de forma dinâmica. Isso é crucial para capturar a sazonalidade e a interação entre capturas no mesmo local, que podem estar correlacionadas devido a fatores ambientais locais. O AG lida eficientemente com censura intervalar, comum neste estudo devido às visitas periódicas.

Realizou-se a modelagem considerando nos estratos: (i) apenas as primeira capturas e (ii) recapturas.

4 RESULTADOS E DISCUSSÕES

Nesta seção, são apresentados os resultados obtidos com a aplicação da análise de sobrevivência aos dados de captura de enxames de abelhas africanizadas, acompanhados de discussões sobre suas interpretações e implicações.

A caracterização dos ninhos-armadilha e a estruturação dos dados de captura estabeleceram a base para identificar padrões temporais e geográficos.

O modelo de Cox, ajustado com variáveis lineares e circulares, permitiu avaliar os fatores que influenciam o tempo até a captura. A inclusão da covariável época, revelou a influência da sazonalidade nos eventos de captura, destacando a ocorrência de multimodalidade.

4.1 Caracterização do ninho-armadilha e estruturação dos dados de captura

Para cada ninho-armadilha (id), foram registradas as suas características de instalação, incluindo a cor de revestimento do ninho, que podia ser amarelo, azul ou preto; a altura em relação ao solo; a altitude do ninho armadilha, também em metros; a latitude e a longitude, ambas em graus; e o ângulo de posicionamento do alvado em relação ao norte magnético, medido em graus. Esses dados foram registrados conforme apresentado na Tabela 4.1.

Tabela 4.1 – Características físicas e geográficas dos ninhos-armadilha.

id	Cor	Altitude (m)	Altura (m)	Alvado (°)	Latitude (°)	Longitude (°)
1	Azul	924	3,98	311	-21,10635278	-44,24906667
2	Amarelo	924	2,68	69	-21,10633889	-44,24904444
⋮						
11	Preto	920	2,59	271	-21,10635833	-44,24878611
⋮						
39	Preto	892	1,55	123	-21,10347222	-44,25136111
⋮						
51	Preto	943	1,22	312	-21,10661111	-44,25086111
⋮						

Fonte: Do autor (2025).

O posicionamento dos ninhos foram utilizados para incorporar as características de localização para auxiliar na modelagem do evento de interesse, tempo até a captura. No entanto, houve casos com ausência de capturas ou problemas adversos como desgaste das caixas ou de uso por outro inseto, sendo assim necessário o estudo com inclusão de censura. Houve também casos em que em uma localidade, foram capturados mais de um enxame, caracterizado evento recorrente.

Dessa forma, ao longo de 365 dias, informações relacionadas ao estudo de sobrevivência foram adicionadas ao banco de dados, apresentados na Tabela 4.2. O *Estrato* refere-se às reposições dos ninhos devido a captura ou censuras por motivos adversos, sendo 1 para o primeiro ninho instalado e 2 para o ninho que foi repostado no lugar do primeiro e assim sucessivamente. O intervalo de tempo, expresso em dias possíveis de ocorrência de enxameação na armadilha, foi definido pelo *Início*, e pelo *Fim*, que indica o dia da observação da captura, caracterizando uma censura intervalar. O status da ocorrência da captura foi categorizado como 1, quando um enxame de abelhas africanizadas foi observado dentro do intervalo de tempo, ou como 0, quando o ninho permaneceu vazio no intervalo, ou seja, instalado no local e disponível para o evento de interesse (sob risco).

Assim, por exemplo, no ninho-armadilha 1, observa-se que, entre os dias 1 e 151, não houve captura, no intervalo seguinte, entre os dias 152 e 181, uma captura foi registrada. O ninho foi substituído no dia seguinte após a observação da captura, iniciando-se o estrato 2, entre os dias 182 e 365, em que voltou a estar sob risco, sem nova ocupação até o fim do período. Já no ninho 2 não houve captura durante todo o experimento.

No caso do ninho 11, destaca-se a ocorrência de capturas recorrentes ao longo do ano na mesma localidade, com a necessidade de reposição de quatro ninhos-armadilha, sendo três deles com captura registrada entre os dias 8 e 9, 252 e 257, e 316 e 330, evidenciando uma alta atratividade desse ponto. Observe que na mudança dos estratos, neste caso, apresentou-se o caso em que os ninhos não foram repostos no dia seguinte a observação da captura, isso se deve a fatores, tais como: a condição climática, em que nos dias com chuvas não é possível nem a retirada do ninho com o exame nem a instalação de um novo ninho vazio; ou a disponibilidade do local definitivo em que os ninhos seriam manejados, já que a estrutura permanente é feita de madeira que substitui a temporária de caixas arquivos de papelão. Nessas situações o intervalo de tempo não foi considerado, já que os ninhos estavam com enxames e ficaram fora tanto do intervalo de captura quanto do intervalo sob risco para a modelagem de análise de sobrevivência.

Tabela 4.2 – Dados de identificação, períodos de observação e status de captura dos ninhos-armadilha.

id	Estrato	Início	Fim	Captura
1	1	1	151	0
1	1	152	181	1
1	2	182	365	0
2	1	1	365	0
⋮				
11	1	1	7	0
11	1	8	9	1
11	2	11	251	0
11	2	252	257	1
11	3	269	315	0
11	3	316	330	1
11	4	346	365	0
⋮				
39	1	5	56	0
39	2	140	365	0
⋮				
51	1	6	12	0
51	1	13	14	1
51	2	15	78	0
⋮				

Fonte: Do autor (2025).

O ninho 39 ilustra uma situação de censura adversa em que houve ocupação por uma espécie de abelha sem ferrão. Essas espécies estão sujeitas a regulamentação segundo a Lei Federal nº 14.639/2023 (Brasil, 2023) e a Resolução CONAMA nº 456/2020 (CONAMA, 2020). Assim, sua transferência necessita de um período de aproximadamente 90 dias para se estabelecerem e serem remanejadas ao meliponário, mantendo o ninho indisponível para novas capturas de abelhas africanizadas durante esse período. Devido a isso, foi registrado somente o tempo em que o ninho esteve disponível tanto para o estrato 1 quanto para o 2.

O ninho 51 exemplifica outra situação aleatória, em que o ninho-armadilha foi afetado por um evento externo, a queda de uma árvore, conforme ilustrado na Figura 4.1(a). Inicialmente, entre os dias 6 e 12, o ninho estava sob risco. Entre os dias 13 e 14, houve a captura de um enxame de abelhas africanizadas. Após essa captura, o ninho foi substituído e voltou a estar sob risco entre os dias 15 e 78, o último dia em que foi observado na árvore.

Figura 4.1 – Fotografias: (a) Árvore caída com o ninho-armadilha no chão; (b) Enxame de abelha africanizada presente em um ninho-armadilha; (c) Ninho-armadilha mostrando altura e ângulo do alvado.



Fonte: Do autor (2025).

Já a Figura 4.1(b) exibe um exemplo de captura de um enxame de abelhas africanizadas com a marcação de número de caixa exibindo também o número do estrato. Por fim a Figura 4.1(c) ilustra a metodologia empregada para a medição da altura e do ângulo do alvado dos ninhos-armadilha.

4.2 Ajuste do Modelo de Cox

A análise de modelos de sobrevivência foi realizada utilizando o modelo de Cox para avaliar a atratividade de ninhos-armadilha no *Campus* Tancredo de Almeida Neves (CTAN), com a estatística z para teste de significância. Os modelos consideram variáveis como cor do ninho, altura, altitude, ângulo do alvado e localização geográfica, representada pelo seno da latitude e cosseno da longitude normalizados.

4.2.1 Modelos sem recaptura

O primeiro modelo avalia a atratividade dos ninhos considerando todas as cores disponíveis e utiliza o seno da latitude e o cosseno da longitude normalizados pela média para a localização geográfica. Ele foi ajustado com um estrato inicial, analisando 120 observações com 28 eventos.

Tabela 4.3 – Resultados do modelo de Cox completo sem recapturas (n=120, eventos=28).

Covariável	Coeficiente	exp(Coef)	Erro Padrão	z	p-valor
Cor (Preto)	2,319	10,166	0,763	3,040	0,00236 **
Cor (Azul)	2,174	8,792	0,777	2,799	0,00512 **
Altura	0,389	1,476	0,266	1,463	0,14343
Altitude	0,049	1,050	0,041	1,189	0,23428
sen(Alvado)	-0,041	0,960	0,272	-0,150	0,88101
cos(Alvado)	0,231	1,260	0,304	0,759	0,44765
Latitude	0,607	1,834	0,703	0,863	0,38828
Longitude	0,683	1,979	0,293	2,329	0,01984 *

Fonte: Do autor (2025).

O modelo com todas as cores (Tabela 4.3) mostra que os ninhos pretos ($p = 0,00236$) e azuis ($p = 0,00512$) têm riscos de captura significativamente maiores que os amarelos (10,166 e 8,792 vezes, respectivamente). O cosseno da longitude normalizado é significativo ($p = 0,01984$), sugerindo maior atratividade em posições mais a leste da área experimental. Outras variáveis, como altura, altitude e ângulo do alvado, não foram significativas.

Devido a ocorrência de baixa captura nos ninhos de cor amarela, as mesmas foram retiradas do segundo modelo, comparando apenas os ninhos de revestimentos pretos e azuis, utilizou-se o revestimento preto como referência, mantendo o seno da latitude e o cosseno da longitude normalizados. Assim, o novo modelo considerou 87 observações com 26 eventos no estrato inicial.

No modelo sem considerar o ninho com revestimento amarelo (Tabela 4.4) não obteve-se diferença significativa entre as cores preta e azul ($p = 0,87460$), com risco relativo próximo de 1 (0,932), indicando atratividade semelhante no estrato inicial. A longitude mantém significância ($p = 0,03460$), reforçando a relevância da posição leste-oeste.

Tabela 4.4 – Resultados do modelo de Cox completo sem recapturas e sem ninhos revestidos em amarelo (n=87, eventos=26).

Covariável	Coeficiente	exp(Coef)	Erro Padrão	z	p-valor
Cor (Azul)	-0,071	0,932	0,449	-0,158	0,87460
Altura	0,276	1,318	0,245	1,128	0,25940
Altitude	0,042	1,043	0,042	1,017	0,30910
sen(Alvado)	-0,051	0,950	0,293	-0,175	0,86110
cos(Alvado)	0,419	1,520	0,297	1,410	0,15870
Latitude	0,414	1,513	0,696	0,596	0,55140
Longitude	0,658	1,930	0,311	2,113	0,03460 *

Fonte: Do autor (2025).

A análise anterior foi novamente realizada considerando a inclusão da variável época, representada por seno e cosseno.

Tabela 4.5 – Resultados do modelo de Cox completo sem recapturas, sem ninhos revestidos em amarelo, com a inclusão da covariável época (n=87, eventos=26).

Covariável	Coeficiente	exp(Coef)	Erro Padrão	z	p-valor
Cor (Azul)	-0,321	0,725	0,453	-0,708	0,47910
Altura	-0,155	0,857	0,308	-0,503	0,61480
Altitude	0,051	1,052	0,045	1,127	0,25950
sen(Alvado)	-0,056	0,946	0,306	-0,183	0,85480
cos(Alvado)	0,313	1,368	0,283	1,107	0,26820
Latitude	0,689	1,991	0,745	0,924	0,35540
Longitude	0,633	1,883	0,318	1,989	0,04670 *
sen(Época)	0,626	1,870	0,444	1,411	0,15820
cos(Época)	0,874	2,395	0,403	2,170	0,03000 *

Fonte: Do autor (2025).

O modelo com inclusão da época (Tabela 4.5) não identifica diferença significativa entre as cores preta e azul ($p = 0,47910$). O cosseno da época é significativo ($p = 0,03000$), sugerindo influência sazonal na captura, e a longitude mantém significância ($p = 0,04670$). Outras variáveis não apresentam impacto relevante.

4.2.2 Modelos com recapturas

Ao se ajustar o primeiro modelo (AG) considerando o banco de dados completo com recapturas, avaliou-se a atratividade dos ninhos com todas as cores, incluindo seno da latitude e cosseno da longitude normalizados, e considera todos os estratos. Analisou-se 167 observações com 40 eventos nesse caso.

Tabela 4.6 – Resultados do modelo AG completo (n=167, eventos=40).

Covariável	Coefficiente	exp(Coef)	Erro Padrão	z	p-valor
Cor (Preto)	2,513	12,338	0,734	3,423	0,00062 ***
Cor (Azul)	2,018	7,525	0,745	2,710	0,00673 **
Altura	0,220	1,245	0,187	1,176	0,23946
Altitude	0,017	1,017	0,029	0,570	0,56849
sen(Alvado)	-0,231	0,793	0,171	-1,352	0,17624
cos(Alvado)	0,308	1,360	0,288	1,068	0,28563
Latitude	-0,054	0,948	0,544	-0,099	0,92108
Longitude	0,480	1,617	0,244	1,966	0,04928 *

Fonte: Do autor (2025).

O modelo com todas as cores e estratos (Tabela 4.6) confirma a forte influência da cor, com ninhos pretos ($p = 0,00062$) e azuis ($p = 0,00673$) apresentando riscos significativamente maiores que os amarelos (12,338 e 7,525 vezes, respectivamente). A longitude é significativa ($p = 0,04928$), enquanto outras variáveis não mostram significância estatística.

No segundo modelo AG excluiu-se a cor amarelo, comparando apenas os ninhos pretos e azuis, com preto como referência, e considerando todos os estratos com seno da latitude e cosseno da longitude normalizados. Neste caso analisou-se 132 observações com 38 eventos de captura.

No modelo sem a cor amarela e com todos os estratos (Tabela 4.7) não se obteve diferença significativa entre as cores de revestimento preta e azul ($p = 0,16030$), com risco relativo de 0,632. A longitude mostra não significância ($p = 0,07470$), porém, pelo valor p ser menor que 10%, pode haver alguma influência da posição leste-oeste.

No terceiro modelo AG adicionou-se a variável época, analisando 132 observações com 38 eventos de captura.

Tabela 4.7 – Resultados do modelo AG completo sem os ninhos revestidos em amarelo (n=132, eventos=38).

Covariável	Coefficiente	exp(Coef)	Erro Padrão	z	p-valor
Cor (Azul)	-0,459	0,632	0,327	-1,404	0,16030
Altura	0,157	1,170	0,168	0,934	0,35040
Altitude	0,013	1,013	0,029	0,428	0,66830
sen(Alvado)	-0,233	0,792	0,182	-1,280	0,20060
cos(Alvado)	0,426	1,531	0,286	1,488	0,13670
Latitude	-0,186	0,830	0,538	-0,345	0,73000
Longitude	0,462	1,588	0,259	1,783	0,07470

Fonte: Do autor (2025).

Tabela 4.8 – Resultados do modelo com seno/cosseno da época, sem amarelo, seno da latitude e cosseno da longitude normalizados, todos os estratos (n=132, eventos=38).

Covariável	Coefficiente	exp(Coef)	Erro Padrão	z	p-valor
Cor (Azul)	-0,561	0,570	0,278	-2,016	0,04382 *
Altura	-0,136	0,873	0,180	-0,754	0,45100
Altitude	0,016	1,016	0,029	0,562	0,57390
sen(Alvado)	-0,115	0,891	0,189	-0,608	0,54296
cos(Alvado)	0,196	1,217	0,258	0,762	0,44632
Latitude	-0,012	0,988	0,472	-0,025	0,97991
Longitude	0,443	1,557	0,253	1,751	0,07989
sin(Época)	1,020	2,773	0,283	3,603	0,00032 ***
cos(Época)	0,656	1,928	0,357	1,840	0,06576

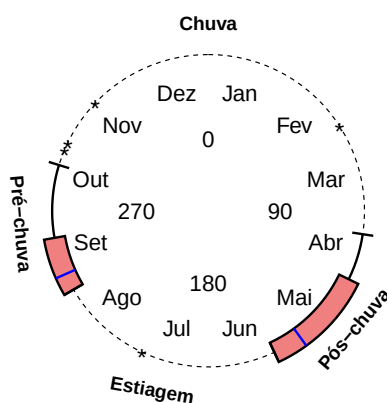
Fonte: Do autor (2025).

O modelo com inclusão da época e todos os estratos (Tabela 4.8) indica que a cor azul é significativamente menos atrativa que a preta ($p = 0,04382$), com risco relativo de 0,570. O seno da época é altamente significativo ($p = 0,00032$), evidenciando forte influência sazonal, enquanto o cosseno da época ($p = 0,06576$) e a longitude ($p = 0,07989$) mostram não significância, porém valor p menor que 10%. Esses resultados sugerem que ninhos pretos são mais eficazes em períodos sazonais específicos.

4.3 Modelos bimodais

Realizou-se o estudo do efeito sazonal, verificando a ocorrência de multimodalidade sendo necessário avaliar o banco de dados em épocas distintas. Utilizou-se para tal estudo o Método de Segmentação do Boxplot Circular conforme metodologia apresentada na seção ?? do Capítulo 1, optando-se por uma segmentação em 2 grupos, conforme discutido nas seções 3 e 4 no Capítulo 2, que resultou na figura 4.2 que apresenta os dois períodos de capturas de enxames de abelhas africanizadas.

Figura 4.2 – Boxplot circular representando a segmentação das épocas de capturas de enxames de abelhas africanizadas, com indicação dos períodos sazonais (Chuva, Pós-chuva, Pré-chuva e Estiagem).



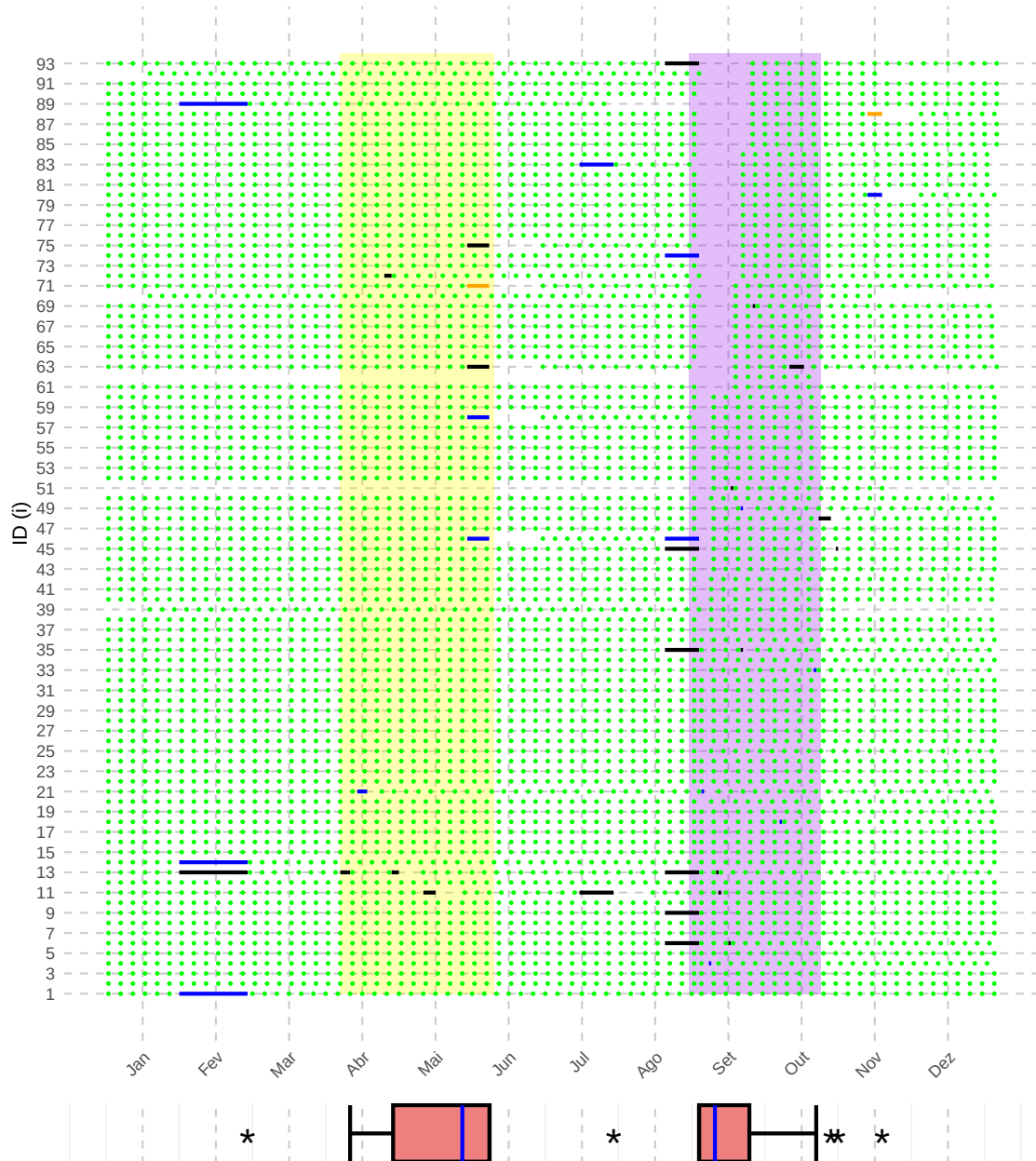
Fonte: Do autor (2025).

O MSBC foi aplicado para dividir o banco de dados em dois, e para cada um deles foi calculado um boxplot. Os boxplots foram nomeados de região de captura pós-chuva, para as capturas no primeiro semestre, e região de captura pré-chuva para o boxplot do segundo semestre. Assim, o conjunto de dados foi segmentado a partir do limites de cada boxplot, como mostra a Figura 4.3.

A representação visual dos dados de sobrevivência foi apresentada de acordo com a identificação do ninho-armadilha e a época do ano, da seguinte maneira:

- linha pontilhada verde: período em que os 93 ninhos estiveram sob risco;
- segmento de reta ou ponto: tempo no qual o enxame capturado permaneceu no ninho-armadilha, de acordo com a cor de revestimento (amarelo, azul ou preto);
- faixa vertical amarela: duração do primeiro período de captura pós-chuva;
- faixa vertical roxa: duração do segundo período de captura pré-chuva.

Figura 4.3 – Gráfico combinado mostrando os intervalos de captura e censura dos ninhos-armadilhas ao longo do ano (parte superior) e boxplot circular linearizado dos períodos de captura segmentados (parte inferior). As cores dos segmentos indicam a cor do ninho-armadilha (azul, preto, amarelo), e os pontos verdes representam observações de censura.



Fonte: Do autor (2025).

O boxplot circular bimodal foi linearizado e posicionado logo abaixo para representar a delimitação dos períodos de capturas, bem como as capturas consideradas como *outliers*.

Na segmentação temporal das capturas de enxames, o primeiro período de captura, região pós-chuva, registrou 10 capturas entre 10 de abril e 6 de junho, e no período da região pré-chuva houve o dobro de capturas entre 29 de agosto e 19 de outubro. A amplitude interquartílica dos boxplots mostram que mesmo em menor quantidade, as capturas do primeiro semestre estão dispersas em um intervalo de 40 dias, já as capturas do segundo semestre estão concentradas em um período de 21 dias.

As primeiras capturas do ano, em fevereiro, são classificadas como *outliers*, e se justifica por um período de estiagem em meio a chuvas consecutivas que duraram meses, mas não se recomenda tentar realizar capturas nesse período. É importante que a instalação de ninhos-armadilha seja feito após o período de chuvas, principalmente se forem fabricadas de papelão, ainda que com uma camada protetora de revestimento é provável que elas se danifiquem nos primeiros dias em caso de dias consecutivos de precipitação.

Para realizar capturas no primeiro semestre os ninhos armadilhas devem estar instaladas até o início de abril, com expectativa de enxameação até o início de junho. Por ser um período seco não deve ocorrer a danificação dos ninhos por conta de chuvas, e portanto não há necessidade de retirada dos mesmos. Podendo ocorrer capturas esporádicas nas semanas seguintes ao fim do período de captura, como os *outliers* registrados em julho.

O principal período de captura se inicia na segunda metade de agosto, até o fim de setembro, um período de 45 dias responsável por 50% das capturas do experimento. Caso haja necessidade de remoção das caixas para reaproveitá-las para o ano seguinte é importante se certificar que elas sejam retiradas antes que inicie o período chuvoso, sendo então outubro um mês de alerta, que tem uma probabilidade moderada de acontecer capturas até a terceira semana, mas deve-se levar em conta o risco de perder o ninho-armadilha definitivamente. Mas deve-se registrar que o melhor período para instalar ninhos armadilhas é na metade final de agosto, e espera-se que 50% das capturas desse período já tenham ocorrido nos primeiros dias de setembro, com expectativa de mais capturas até o fim do mês.

Outra orientação válida é a reinstalação de ninhos nos locais com capturas pelo menos até o fim de setembro, para outubro não é recomendado a reinstalação, mesmo que possa ocorrer mais algumas capturas com o início do período chuvoso é provável que as caixas se danifiquem. Que são os casos dos *outliers* registrados no fim de outubro, são capturas que poderiam pertenc-

cer ao grupo de capturas do segundo semestre, porém o fim de agosto e início de setembro tiveram uma concentração tão alta de capturas que a amplitude interquartílica ficou reduzida.

Para identificar os fatores mais relevantes que influenciam o tempo até a captura de enxames de abelhas africanizadas, foi aplicada a técnica de seleção *stepwise* com critério de informação de Akaike (AIC) aos modelos de Andersen-Gill (AG) ajustados para os períodos pós-chuva (Grupo 1) e pré-chuva (Grupo 2). Essa abordagem combina inclusão progressiva e exclusão regressiva de covariáveis, buscando um modelo parcimonioso que equilibre ajuste e simplicidade. Os modelos finais foram validados por meio de testes de proporcionalidade dos riscos e análise de resíduos, garantindo sua robustez. A função de risco para cada modelo foi expressa na forma $\lambda(t_{ij} | x_i) = \lambda_0(t_{ij}) \exp(y)$, onde y representa a combinação linear das covariáveis selecionadas.

4.3.1 Período Pós-chuva

A Tabela 4.9 apresenta os resultados do modelo de Andersen-Gill aplicado ao período pós-chuva, com 105 observações e 10 eventos de captura.

Tabela 4.9 – Resultados do modelo AG completo para o período pós-chuva (n=105, eventos=10).

Covariável	Coefficiente	exp(Coef)	Erro Padrão	z	p-valor
Cor (Preto)	2,104	8,198	0,980	2,147	0,03180 *
Cor (Azul)	0,958	2,605	1,178	0,813	0,41610
Altura	0,544	1,722	0,442	1,231	0,21850
Altitude	0,023	1,023	0,070	0,324	0,74600
sen(Alvado)	-1,003	0,367	0,361	-2,779	0,00540 **
cos(Alvado)	0,691	1,995	0,571	1,210	0,22630
Latitude	0,470	1,599	1,246	0,377	0,70620
Longitude	0,225	1,253	0,310	0,727	0,46750

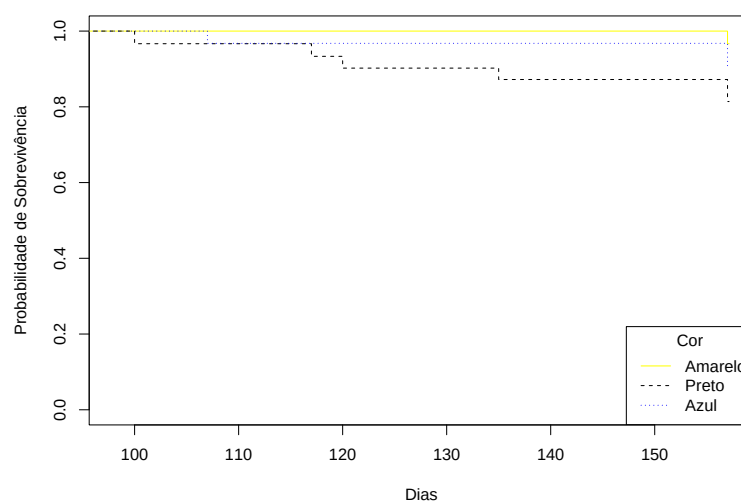
Fonte: Do autor (2025).

Este período, caracterizado por baixas temperaturas após o período chuvoso, reflete um momento em que as abelhas africanizadas, espécie de sangue frio, são mais rigorosas na escolha de locais para nidificação, buscando ninhos que ofereçam condições térmicas estáveis e proteção contra variações climáticas. Segundo Stabentheiner, Kovac e Brodschneider (2010)

larvas e pupas de abelhas são extremamente estenotérmicas, ou seja, dependem fortemente da regulação precisa da temperatura do ninho de cria para seu desenvolvimento adequado (33–36 °C).

A cor preta do ninho-armadilha é significativamente mais atrativa ($p=0,0318$), uma justificativa pode ser a maior absorção de radiação solar em comparação com cores mais claras, com um risco relativo de 8,198 vezes maior que o amarelo, enquanto a cor azul não apresentou diferença significativa ($p=0,4161$), como ilustrado na curva de sobrevivência da Figura 4.4.

Figura 4.4 – Curva de sobrevivência estimada para o tempo até a captura de enxames de abelhas africanizadas, por cor de revestimento do ninho-armadilha, no período pós-chuva.



Fonte: Do autor (2025).

O seno do alvado é altamente significativo ($p=0,0054$), indicando que a orientação da entrada do ninho é um fator crítico na escolha das abelhas, particularmente em relação à incidência de luz solar na caixa. Este resultado corrobora com os estudos de Meikle, Weiss e and (2023) que posicionaram colmeias com entradas voltadas para os quatro pontos cardeais e monitoraram o peso e a temperatura interna das colmeias ao longo de 14 meses. Os resultados indicaram que a orientação da entrada da colmeia pode influenciar significativamente a atividade das abelhas, especialmente em climas mais frios.

Essa preferência sugere que as abelhas buscam ângulos de alvado que maximizem a captação de luz e calor solar, essenciais para manter a temperatura interna da colmeia em períodos mais frios. A ausência de significância para altura, altitude, latitude e longitude reforça que, em

condições de temperaturas mais baixas, as abelhas priorizam características do ninho, como cor e orientação do alvado para otimizar a incidência solar.

O processo *stepwise* iniciou-se com o modelo completo (AIC = 93,67). Após sucessivas iterações, foram removidas, nesta ordem, Altitude, cos(Longitude), sin(Latitude), Altura e cos(Alvado). A função de risco do modelo final (AIC = 87,62) é dada por:

$$\lambda(t_{ij} | x_i) = \lambda_0(t_{ij}) \exp(1,9 \cdot \text{Preto} + 1,05 \cdot \text{Azul} - 1,03 \cdot \sin(\text{Alvado})).$$

O modelo indica que ninhos com revestimento preto apresentam risco de captura 6,72 vezes maior que os amarelos ($p = 0,0781$), enquanto a cor azul não foi significativa ($p = 0,3541$). A covariável sin(Alvado) ($p = 0,0194$) sugere que a orientação do alvado é um fator crítico, com coeficiente negativo indicando preferência por ângulos que maximizam a incidência solar. A suposição de riscos proporcionais foi validada pelo teste de Schoenfeld (p -valor global = 0,73), com p -valores individuais de 0,53 para Cor e 0,88 para sin(Alvado).

4.3.2 Período Pré-chuva

A Tabela 4.10 apresenta os resultados do modelo AG aplicado ao período pré-chuva, com 205 observações e 20 eventos de captura.

Tabela 4.10 – Resultados do modelo AG completo para o período pré-chuva (n=205, eventos=20).

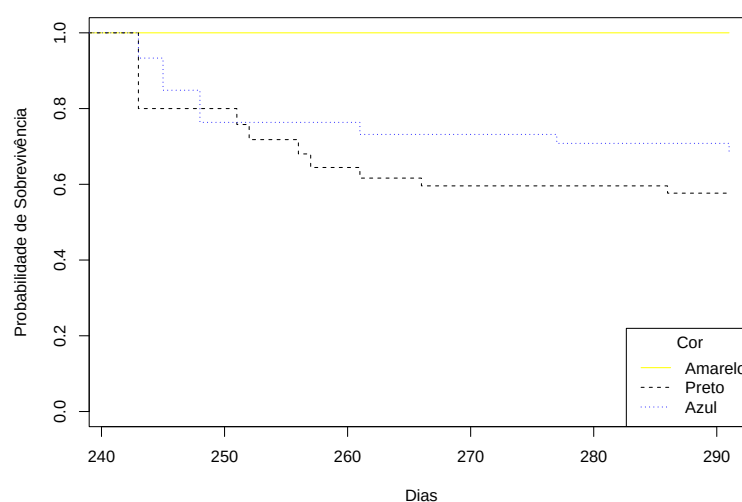
Covariável	Coeficiente	exp(Coef)	Erro Padrão	z	p-valor
Cor (Preto)	20,100	5,365e+08	0,370	54,294	<0,00001 ***
Cor (Azul)	19,430	2,750e+08	0,389	49,972	<0,00001 ***
Altura	0,030	1,030	0,257	0,115	0,90900
Altitude	0,038	1,038	0,043	0,880	0,37900
sen(Alvado)	-0,208	0,812	0,284	-0,731	0,46500
cos(Alvado)	0,213	1,237	0,327	0,651	0,51500
Latitude	0,127	1,135	0,759	0,167	0,86700
Longitude	0,424	1,529	0,400	1,060	0,28900

Fonte: Do autor (2025).

Este período, caracterizado por temperaturas mais elevadas e antecedendo o início do período chuvoso, reflete um momento de alta atividade de enxameação das abelhas africanas.

zadas, quando as condições climáticas favorecem a busca por novos locais de nidificação. A maior quantidade de capturas neste período, em comparação com o período pós-chuva, sugere que fatores climáticos como discutido na seção 3 do Capítulo 2 e a consequente disponibilidade de recursos florais.

Figura 4.5 – Curva de sobrevivência estimada para o tempo até a captura de enxames de abelhas africanizadas, por cor de revestimento do ninho-armadilha, no período pré-chuva.



Fonte: Do autor (2025).

A incorporação da variável cor foi super valorizada em função da não ocorrência de capturas de ninho de revestimento amarelo, indicando que as cores preta com 13 capturas ($p < 2e - 16$) e azul com 7 ($p < 2e - 16$) são extremamente atrativas em comparação com a cor amarela, como ilustrado na curva de sobrevivência da Figura 4.5. A ausência de significância para altura, altitude, orientação do alvado, latitude e longitude indica que, em condições pré-chuva, as abelhas africanizadas tem alguma restrição com relação as caixas amarelas.

O processo *stepwise* iniciou-se com o modelo completo (AIC = 162,41). Após sucessivas iterações, foram removidas, nesta ordem, *Altura*, $\sin(\text{latitude})$, $\sin(\text{Alvado})$ e $\cos(\text{Alvado})$. A função de risco do modelo final (AIC = 155,40) é dada por:

$$\lambda(t_{ij} | x_i) = \lambda_0(t_{ij}) \exp(20,1 \cdot \text{Preto} + 19,43 \cdot \text{Azul} + 0,03 \cdot \text{Altitude} + 0,41 \cdot \cos(\text{Longitude})).$$

O modelo final confirma a alta atratividade dos ninhos pretos e azuis ($p < 2 \times 10^{-16}$), com riscos relativos extremamente elevados em comparação com ninhos amarelos, que não registraram capturas. A covariável $\cos(\text{longitude})$ ($p = 0,1427$) indica um efeito não significativo, mas com possível influência da posição leste-oeste. A suposição de riscos proporcionais

foi validada pelo teste de Schoenfeld (p-valor global = 0,15), com p-valores individuais de Cor = 0,68, Altitude = 0,14, $\cos(\text{Longitude}) = 0,25$.

5 CONCLUSÕES

A análise de sobrevivência aplicada ao estudo do tempo até a captura de enxames de abelhas africanizadas em ninhos-armadilhas revelou fatores que influenciam a atratividade das colmeias, contribuindo para a otimização das práticas apícolas.

Utilizando o modelo de Andersen-Gill (AG) para eventos recorrentes e integrando estatística circular, foi possível modelar a interação entre variáveis lineares (cor, altura, altitude) e circulares (orientação do alvado, localização geográfica, sazonalidade), capturando a natureza cíclica de variáveis angulares.

Os resultados destacaram a cor do revestimento como fator determinante. Ninhos com revestimento nas cores pretas e azuis apresentaram riscos de captura significativamente maiores que os amarelos.

Altura, altitude e latitude não apresentaram significância neste estudo, no entanto, a posição geográfica que se registra na longitude apresentou significância quando se avalia a captura ocorrida anualmente, provavelmente devido as mudanças de posicionamento em relação ao Sol. Além disso, a inclusão da covariável Época do ano destacou a influência sazonal nas capturas, com períodos específicos sendo mais favoráveis para a captura de enxames.

O modelo de Cox foi utilizado para avaliar esses fatores, e a sazonalidade também foi considerada, revelando padrões de captura em diferentes épocas do ano, evidenciando a bimodalidade do evento de interesse.

A Época do Ano, analisada pelo Método de Segmentação do Boxplot Circular, confirmou dois períodos de captura: o pós-chuva, com capturas dispersas, e o pré-chuva, responsável por 50% das capturas, concentradas entre o final de agosto e início de setembro.

No período pós-chuva (abril a junho) o revestimento dos ninhos-armadilha na cor preta mostrou-se mais eficaz, possivelmente devido à maior absorção de radiação solar, favorecendo a regulação térmica em temperaturas mais baixas. No período pré-chuva (agosto a outubro), ambas as cores azul e preta foram altamente atrativas, indicando menor seletividade das abelhas em condições de alta atividade de enxameação, exceto pelo amarelo, que mostrou baixa eficiência.

A orientação do alvado foi significativa no período pós-chuva, sugerindo que as abelhas preferem ângulos que maximizam a incidência solar para manter a temperatura interna da colmeia.

Assim, recomenda-se na captura de enxames usar ninhos pretos ou azuis, evitando amarelos; instalar ninhos no início de abril para o período pós-chuva e no final de agosto para o pré-chuva, mantendo-os até o final de setembro; ajustar a orientação do alvado no período pós-chuva para otimizar a captação de luz solar; repor ninhos-armadilha em locais com ocorrência de capturas até o final de setembro, evitando outubro devido ao risco de danos por chuvas.

REFERÊNCIAS

- AKKUZU, E. *et al.* Assesment of trap color and trap height above the ground on the capture of *ips sexdentatus* and *thanasimus formicarius*. **Šumar. List**, v. 145, p. 169–174, 2021.
- ANDERSEN, P. K.; GILL, R. D. Cox's regression model for counting processes: a large sample study. **The Annals of Statistics**, Hayward, v. 10, n. 4, p. 1100–1120, Dec 1982.
- ANDERSEN, P. K. *et al.* Analysis of time-to-event for observational studies: guidance to the use of intensity models. **Statistics in medicine**, Wiley Online Library, v. 40, n. 1, p. 185–211, 2021.
- ANTOINE, C. M.; FORREST, J. R. Nesting habitat of ground-nesting bees: a review. **Ecological Entomology**, Wiley Online Library, v. 46, n. 2, p. 143–159, 2021.
- Brasil. **Lei Federal nº 14.639, de 25 de julho de 2023**. 2023. Dispõe sobre a proteção e o incentivo à criação e à preservação de abelhas nativas sem ferrão. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/lei/L14639.htm.
- BRODSCHNEIDER, R.; GRAY, A. How coloss monitoring and research on lost honey bee colonies can support colony survival. **Bee World**, Taylor & Francis, v. 99, n. 1, p. 8–10, 2022.
- CARVALHO, M. S. *et al.* **Análise de sobrevivência: teoria e aplicações em saúde**. [S.l.]: SciELO-Editora FIOCRUZ, 2011.
- CÉSARD, N. An emic understanding of honey bees and their environment: attracting bee swarms to nest on rafters in belitung, indonesia. **Anthropozoologica**, BioOne, v. 57, n. 10, p. 223–239, 2022.
- CHIOU, S. H. *et al.* Regression modeling for recurrent events possibly with an informative terminal event using r package rereg. **Journal of Statistical Software**, v. 105, p. 1–34, 2023.
- CHUNG, Y. *et al.* Partial likelihood estimation of isotonic proportional hazards models. **Biometrika**, Oxford University Press, v. 105, n. 1, p. 133–148, 2018.
- CONAMA. **Resolução CONAMA nº 496, de 19 de agosto de 2020**. 2020. Estabelece diretrizes para o manejo e conservação de abelhas nativas. Disponível em: <https://www.in.gov.br/en/web/dou/-/resolucao-n-496-de-19-de-agosto-de-2020-273217120>.
- COX, D. R. Regression models and life-tables: with discussions. **Journal of the Royal Statistical Society B: Statistical**, London, v. 34, p. 187–220, 1972.
- COX, D. R.; OAKES, D. **Analysis of survival data**. London: Chapman and Hall, 1984.
- FAROOQ, Z. *et al.* Impact of climate and aedes albopictus establishment on dengue and chikungunya outbreaks in europe: a time-to-event analysis. **The Lancet Planetary Health**, Elsevier, v. 9, n. 5, p. e374–e383, 2025.
- IJIMA, T.; MOMOZAKI, T.; ANDO, S. On window mean survival time with interval-censored data. **arXiv preprint arXiv:2502.20011**, 2025.
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, London, v. 53, n. 282, p. 457–481, 1958.

- KELLY, P. J.; LIM, L. L. Y. Survival analysis for recurrent event data: an application to childhood infectious diseases. **Statistics in Medicine**, New York, v. 19, n. 1, p. 13–33, 2000.
- LAMBRECHTS, M. M.; DEEMING, D. C. Nest design and breeding success: replicability of methodologies and research findings in secondary hole nesting passerines. **Birds**, MDPI, v. 5, n. 2, p. 278–307, 2024.
- LAWLESS, J. F. **Statistical models and methods for lifetime data**. New York: Wiley, 1982.
- LEVEN, L. *et al.* A apicultura nas regiões tropicais. **Série Agrodok No. 32**, Agromisa, 2006.
- LIANG, J. *et al.* Improve individual treatment by comparing treatment benefits: cancer artificial intelligence survival analysis system for cervical carcinoma. **Journal of Translational Medicine**, Springer, v. 20, n. 1, p. 293, 2022.
- MEIKLE, W. G.; WEISS, M.; AND, E. B. Effects of hive entrance orientation on honey bee colony activity. **Journal of Apicultural Research**, Taylor & Francis, v. 62, n. 3, p. 444–449, 2023.
- MONCHARMON, F. X. D. *et al.* Statistical analysis of honeybee survival after chronic exposure to insecticides. **Environmental Toxicology and Chemistry**, New York, v. 22, n. 12, p. 3088–3094, 2003.
- MOUILLARD-LAMPLE, L. *et al.* Competition between wild and honey bees: Floral resources as a common good providing multiple ecosystem services. **Ecosystem Services**, Elsevier, v. 62, p. 101538, 2023.
- RAHIMI, E.; BARGHJELVEH, S.; DONG, P. How effective are artificial nests in attracting bees? a review. **Journal of Ecology and Environment**, Springer, v. 45, p. 1–11, 2021.
- RASHEED, M. Analyzing applications and properties of the exponential continuous distribution in reliability and survival analysis. **Journal of Positive Sciences**, v. 4, n. 5, p. 71–79, 2023.
- RENTERIA, E.; BREHM, G. Is blue the most attractive color for bees? exploring the attractiveness of colors in vane traps. **bioRxiv**, Cold Spring Harbor Laboratory, p. 2025–02, 2025.
- SELINGEROVA, I.; KATINA, S.; HOROVA, I. Comparison of parametric and semiparametric survival regression models with kernel estimation. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 91, n. 13, p. 2717–2739, 2021.
- SIEHLER, O.; WANG, S.; BLOCH, G. Remarkable sensitivity of young honey bee workers to multiple non-photoc, non-thermal, forager cues that synchronize their daily activity rhythms. **Frontiers in Physiology**, Frontiers Media SA, v. 12, p. 789773, 2021.
- SILVA, D. P. d. **Introdução ao modelo de Cox com aplicação a dados de pneus 11.00 r22**. 40 p. Dissertação (Monografia (Graduação em Estatística)) — Universidade Federal da Paraíba, Campina Grande, 2016.

STABENTHEINER, A.; KOVAC, H.; BRODSCHNEIDER, R. Honeybee colony thermoregulation–regulatory mechanisms and contribution of individuals in dependence on age, location and thermal stress. **PLoS one**, Public Library of Science San Francisco, USA, v. 5, n. 1, p. e8967, 2010.

SULTANA, N. *et al.* Evaluating the efficiency of supplementary feeding as a management strategy for enhancing honeybee (*apis mellifera* L.) colony growth and productivity. **Frontiers in Bee Science**, Frontiers Media SA, v. 2, p. 1386799, 2024.

TESFAY, B. G. *et al.* Survival status and predictors of mortality among road traffic accident adult patients admitted to intensive care units of referral hospitals in tigray 2024. **PloS one**, Public Library of Science San Francisco, CA USA, v. 20, n. 3, p. e0308584, 2025.

TOMAZELLA, V. L. D. **Modelagem de dados de eventos recorrentes via processo de Poisson com termo de fragilidade**. 165 p. Tese (Doutorado) — Universidade de São Paulo, São Paulo, 2003.

TURKSON, A. J.; AYIAH-MENSAH, F.; NIMOH, V. Handling censoring and censored data in survival analysis: a standalone systematic literature review. **International journal of mathematics and mathematical sciences**, Wiley Online Library, v. 2021, n. 1, p. 9307475, 2021.

TURNBULL, B. W. The empirical distribution function with arbitrarily grouped, censored and truncated data. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 38, n. 3, p. 290–295, 1976.

VALE, J. O.; ACOSTA-AVALOS, D. Magnetosensitivity in the stingless bee *tetragonisca angustula*: magnetic inclination can alter the choice of the flying departure angle from the nest. **Bioelectromagnetics**, Wiley Online Library, v. 42, n. 1, p. 51–59, 2021.

TERCEIRA PARTE

3 CONSIDERAÇÕES FINAIS

O Método de Segmentação por Boxplot Circular foi apresentado e comparado em simulações, demonstrando equilíbrio entre simplicidade e precisão; suas aplicações em dados de acidentes de trânsito (Artigo 2) e captura de enxames de abelhas (Artigo 3) revelaram padrões multimodais; a integração com análise de sobrevivência no Artigo 4, via modelo Andersen-Gill, modelou eventos recorrentes, identificando fatores influentes como cor e orientação do alvado.

O MSBC em dados reais multimodais confirmaram sua eficácia, com menor rejeição em testes de adequação à von Mises em acidentes e detecção de bimodalidade em capturas de enxames.

O MSBC facilitou segmentação em dados cíclicos, subsidiando políticas de segurança viária ao identificar horários críticos; na apicultura, revelou preferências sazonais e de configuração, contribuindo para a sociedade com práticas sustentáveis, para a ciência com avanço em estatística circular aplicada, e implicações práticas para apicultores, como uso de ninhos pretos/azuis e orientação solar.

Sugere-se pesquisas futuras para integrar a estatística circular à análise de sobrevivência, estudando o tempo como uma variável circular. Modelos paramétricos lineares, como as distribuições exponencial, Weibull e lognormal, são amplamente utilizados; assim, um modelo paramétrico baseado na distribuição von Mises pode servir como ponto de partida promissor, adaptando-se à natureza cíclica do tempo. A multimodalidade representa um desafio particular para a análise de sobrevivência ao tratar o tempo como dado circular, exigindo abordagens inovadoras para lidar com múltiplos picos e dependências temporais.

REFERÊNCIAS

- ANDERSEN, P. K.; GILL, R. D. Cox's regression model for counting processes: a large sample study. **The Annals of Statistics**, Hayward, v. 10, n. 4, p. 1100–1120, Dec 1982.
- ANDERSEN, P. K. *et al.* Analysis of time-to-event for observational studies: guidance to the use of intensity models. **Statistics in medicine**, Wiley Online Library, v. 40, n. 1, p. 185–211, 2021.
- ANDERSON, C. M. **Graphical methods for circular and cylindrical data**. [S.l.], 1993.
- ASUERO, A. G.; SAYAGO, A.; GONZÁLEZ, A. The correlation coefficient: An overview. **Critical reviews in analytical chemistry**, Taylor & Francis, v. 36, n. 1, p. 41–59, 2006.
- BANERJEE, A. *et al.* Clustering on the unit hypersphere using von mises-fisher distributions. **Journal of Machine Learning Research**, v. 6, p. 1345–1382, 2005.
- BEIER, W.; LINDAUER, M. Der sonnenstand als zeitgeber für die biene. **Apidologie**, v. 1, n. 1, p. 5–28, 1970.
- BOUYEYRON, C.; BRUNET-SAUMARD, C. Model-based clustering of high-dimensional data: A review. **Computational Statistics & Data Analysis**, Elsevier, v. 71, p. 52–78, 2014.
- BRODSCHNEIDER, R.; GRAY, A. How coloss monitoring and research on lost honey bee colonies can support colony survival. **Bee World**, Taylor & Francis, v. 99, n. 1, p. 8–10, 2022.
- BRUNSDON, C.; CORCORAN, J. Using circular statistics to analyse time patterns in crime incidence. **Computers, Environment and Urban Systems**, Elsevier, v. 30, n. 3, p. 300–319, 2006.
- BRUZZONE, L.; PRIETO, D. F. An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images. **IEEE Transactions on image processing**, IEEE, v. 11, n. 4, p. 452–466, 2002.
- BUTTARAZZI, D. **Package 'bpDir'**. [S.l.], 2022. R package version 1.0.0.
- BUTTARAZZI, D.; PANDOLFO, G.; PORZIO, G. C. **A boxplot for circular data**. [S.l.]: Oxford, 2018.
- CAPALDI, E. A.; DYER, F. C. The role of orientation flights on homing performance in honeybees. **Journal of Experimental Biology**, v. 202, n. 12, p. 1655–1666, 1999.
- CARVALHO, M. S. *et al.* **Análise de sobrevivência: teoria e aplicações em saúde**. [S.l.]: SciELO-Editora FIOCRUZ, 2011.
- CHAKRABORTY, S.; HAZARIKA, P. J.; ALI, M. M. A multimodal skewed extension of normal distribution: its properties and applications. **Statistics**, Taylor & Francis, v. 49, n. 4, p. 859–877, 2014.
- CHANG-CHIEN, S.-J.; HUNG, W.-L.; YANG, M.-S. On mean shift-based clustering for circular data. **Soft Computing**, Springer, v. 16, p. 1043–1060, 2012.

- CHASANI, P.; LIKAS, A. The uu-test for statistical modeling of unimodal data. **Pattern Recognition**, Elsevier, v. 122, p. 108272, 2022.
- CHASANI, P.; LIKAS, A. Statistical modeling of univariate multimodal data. **arXiv preprint arXiv:2412.15894**, 2024.
- CHIOU, S. H. *et al.* Regression modeling for recurrent events possibly with an informative terminal event using r package rereg. **Journal of Statistical Software**, v. 105, p. 1–34, 2023.
- CHUNG, Y. *et al.* Partial likelihood estimation of isotonic proportional hazards models. **Biometrika**, Oxford University Press, v. 105, n. 1, p. 133–148, 2018.
- COUVILLON, M. J.; SCHÜRCH, R.; RATNIEKS, F. L. Waggle dance distances as integrative indicators of seasonal foraging challenges. **PloS one**, Public Library of Science San Francisco, USA, v. 9, n. 4, p. e93495, 2014.
- COX, D. R. Regression models and life-tables: with discussions. **Journal of the Royal Statistical Society B: Statistical**, London, v. 34, p. 187–220, 1972.
- COX, D. R.; OAKES, D. **Analysis of survival data**. London: Chapman and Hall, 1984.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. **Journal of the royal statistical society: series B (methodological)**, Wiley Online Library, v. 39, n. 1, p. 1–22, 1977.
- DONNER, A.; GOLDENSHLUGER, A. Statistical inference in circular structural model and fitting circles to noisy data. **Journal of Computational and Graphical Statistics**, Taylor & Francis, p. 1–20, 2024.
- ESQUIVEL, D. M. S. *et al.* Do geomagnetic storms change the behaviour of the stingless bee guirucu (schwarziana quadripunctata)? **Naturwissenschaften**, v. 94, p. 139–142, 2007.
- FAROOQ, Z. *et al.* Impact of climate and aedes albopictus establishment on dengue and chikungunya outbreaks in europe: a time-to-event analysis. **The Lancet Planetary Health**, Elsevier, v. 9, n. 5, p. e374–e383, 2025.
- FISHER, N. I.; LEWIS, T.; EMBLETON, B. J. J. **Statistical Analysis of Circular Data**. Cambridge: Cambridge University Press, 1987. ISBN 9780521426726.
- GOULD, J. L.; KIRSCHVINK, J. L.; DEFFEYES, K. S. Bees have magnetic remanence. **Science**, v. 201, n. 4360, p. 1026–1028, 1978.
- HORNIK, K.; LEISCH, F. **movMF: Mixture of von Mises-Fisher Distributions**. [S.l.], 2014. R package version 1.2.0.
- IJIMA, T.; MOMOZAKI, T.; ANDO, S. On window mean survival time with interval-censored data. **arXiv preprint arXiv:2502.20011**, 2025.
- JAMMALAMADAKA, S.; SENGUPTA, A. Circular statistics with applications. **Journal of Statistical Planning and Inference**, Elsevier, v. 99, n. 1, p. 1–20, 2001.
- JAMMALAMADAKA, S. R.; SENGUPTA, A. **Topics in Circular Statistics**. Singapore: World Scientific, 2001. ISBN 9789810242257.

- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, London, v. 53, n. 282, p. 457–481, 1958.
- KELLY, P. J.; LIM, L. L. Y. Survival analysis for recurrent event data: an application to childhood infectious diseases. **Statistics in Medicine**, New York, v. 19, n. 1, p. 13–33, 2000.
- KESEMEN, O.; TEZEL, ; ÖZKUL, E. Fuzzy c-means clustering algorithm for directional data (fcm4dd). **Expert Systems with Applications**, Elsevier, v. 58, p. 76–82, 2016.
- LACERDA, D. C. O. *et al.* Influência dos pontos cardeais e colaterais na nidificação de abelhas nativas em colmeias octogonais. **Gaia Scientia**, v. 11, n. 2, p. 203–217, 2017.
- LANDLER, L.; RUXTON, G. D.; MALKEMPER, E. P. Circular data in biology: advice for effectively implementing statistical procedures. **Behavioral ecology and sociobiology**, Springer, v. 72, p. 1–10, 2018.
- LAWLESS, J. F. **Statistical models and methods for lifetime data**. New York: Wiley, 1982.
- LEVEN, L. *et al.* A apicultura nas regiões tropicais. **Série Agrodok No. 32**, Agromisa, 2006.
- LIANG, J. *et al.* Improve individual treatment by comparing treatment benefits: cancer artificial intelligence survival analysis system for cervical carcinoma. **Journal of Translational Medicine**, Springer, v. 20, n. 1, p. 293, 2022.
- LUCANO, M. J. **Caracterização de material magnético mineralizado em abelhas meliponini *Schwarziana quadripunctata*: magnetometria SQUID e Ressonância Ferromagnética**. 33–37 p. Tese (Doutorado) — Centro Brasileiro de Pesquisas Físicas, 2006.
- LUND, U.; AGOSTINELLI, C.; AGOSTINELLI, M. C. **Package ‘circular’**. [S.l.], 2017. v. 775, n. 5.
- MACEDO, C. R. C. *et al.* Comportamento da nidificação de abelhas melíponas. **Ciência Animal Brasileira**, v. 21, 2020.
- MARDIA, K. Linear-circular correlation coefficients and rhythmometry. **Biometrika**, JSTOR, p. 403–405, 1976.
- MARDIA, K. V. **Statistics of Directional Data**. [S.l.]: Academic Press, 1972.
- MARDIA, K. V.; JUPP, P. E. **Directional Statistics**. [S.l.]: Wiley, 2000.
- MARKOVIĆ, I.; PETROVIĆ, I. Bearing-only tracking with a mixture of von mises distributions. In: IEEE. **2012 IEEE/RSJ International Conference on Intelligent Robots and Systems**. [S.l.], 2012. p. 707–712.
- MEER, L. J. P. van der; B., K. M. **movMF: Mixture of von Mises-Fisher Distributions**. [S.l.], 2021. R package version 1.0.0.
- MISES, R. von. Über die “ganzzahligkeit” der atomgewichte und verwandte fragen. In: SHEN, A. (Ed.). **Selected Papers of Richard von Mises**. [S.l.]: Dover Publications, 1918. p. 58–80.
- MONCHARMON, F. X. D. *et al.* Statistical analysis of honeybee survival after chronic exposure to insecticides. **Environmental Toxicology and Chemistry**, New York, v. 22, n. 12, p. 3088–3094, 2003.

- MORALES, F. **Estimação dos parâmetros de um círculo para modelos heteroscedásticos de regressão**. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2006.
- NEVES, M. R. **Modelos para séries temporais de dados circulares**. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2015.
- PEWSEY, A. The large-sample joint distribution of key circular statistics. **Metrika**, Springer, v. 60, n. 1, p. 25–32, 2004.
- PEWSEY, A.; GARCÍA-PORTUGUÉS, E. Recent advances in directional statistics. **Test**, Springer, v. 30, n. 1, p. 1–58, 2021.
- PEWSEY, A.; NEUHAUS, G.; DAVIS, R. **Circular Statistics: Analysis of Circular Data**. [S.l.]: Wiley, 2013.
- PINTO, F. de A.; SILVA, E. C. da; ANJOS, A. J. dos. Influências climáticas no perfil sanitário em colônias de apis mellifera africanizadas. **Revista Técnica Ciências Ambientais**, v. 1, n. 8, 2024.
- QARMALAH, N. M.; EINBECK, J.; COOLEN, F. P. K-boxplots: An extension of boxplot for circular data. **Statistical Papers**, Springer, v. 57, n. 4, p. 975–990, 2016.
- QU, Y. *et al.* Biology of bamboo nesting stingless bees of southwest china. **Apidologie**, Springer, v. 53, n. 5, p. 53, 2022.
- RASHEED, M. Analyzing applications and properties of the exponential continuous distribution in reliability and survival analysis. **Journal of Positive Sciences**, v. 4, n. 5, p. 71–79, 2023.
- RODRÍGUEZ, C. E.; NÚÑEZ-ANTONIO, G.; ESCARELA, G. A bayesian mixture model for clustering circular data. **Computational Statistics & Data Analysis**, Elsevier, v. 143, p. 106842, 2020.
- SANTOS, W. R. D.; CARNEIRO, T. C. J.; OLIVEIRA, M. P. V. D. Consumidor de dispositivos eletrônicos portáteis: O efeito moderador do perfil do adotante na intenção de compra. **ReMark-Revista Brasileira de Marketing**, v. 15, n. 4, p. 496–511, 2016.
- SELINGEROVA, I.; KATINA, S.; HOROVA, I. Comparison of parametric and semiparametric survival regression models with kernel estimation. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 91, n. 13, p. 2717–2739, 2021.
- SILVA, D. P. d. **Introdução ao modelo de Cox com aplicação a dados de pneus 11.00 r22**. 40 p. Dissertação (Monografia (Graduação em Estatística)) — Universidade Federal da Paraíba, Campina Grande, 2016.
- SILVA, I. S. *et al.* Comparação da produção de abelhas da espécie apis mellifera com as mudanças ambientais. **Animal em Foco**, v. 3, p. 68–76, 2023.
- SOUSA, D. S. R.; RIZZARDO, R. A. G. Dinâmica populacional de colônias de apis mellifera durante o período chuvoso na região de araguaína. **Revista Brasileira de Zootecnia**, Universidade Federal do Tocantins, Araguaína, TO, Brasil, v. 43, n. 5, p. 250–256, 2014.

SOUZA, F. A. M. d. **Influência local e análise de resíduos em modelos de regressão Von Mises**. Tese (Doutorado) — Instituto de Matemática e Estatística da Universidade de São Paulo, 1999.

STEFFAN-DEWENTER, I.; KUHN, A. Honeybee foraging in differentially structured landscapes. **Proceedings of the Royal Society of London. Series B: Biological Sciences**, The Royal Society, v. 270, n. 1515, p. 569–575, 2003.

TEAM, C. R. **R: A Language and Environment for Statistical Computing**. [S.l.], 2025. R Foundation for Statistical Computing, Vienna, Austria.

TESFAY, B. G. *et al.* Survival status and predictors of mortality among road traffic accident adult patients admitted to intensive care units of referral hospitals in tigray 2024. **PloS one**, Public Library of Science San Francisco, CA USA, v. 20, n. 3, p. e0308584, 2025.

TOMAZELLA, V. L. D. **Modelagem de dados de eventos recorrentes via processo de Poisson com termo de fragilidade**. 165 p. Tese (Doutorado) — Universidade de São Paulo, São Paulo, 2003.

TURKSON, A. J.; AYIAH-MENSAH, F.; NIMOH, V. Handling censoring and censored data in survival analysis: a standalone systematic literature review. **International journal of mathematics and mathematical sciences**, Wiley Online Library, v. 2021, n. 1, p. 9307475, 2021.

TURNBULL, B. W. The empirical distribution function with arbitrarily grouped, censored and truncated data. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 38, n. 3, p. 290–295, 1976.

VAZ, M. A. *et al.* Comportamento de nidificação de melipona subnitida (ducke, 1910) e frieseomelitta sp. no seridó oriental do rio grande do norte, brasil. **Research, Society and Development**, v. 10, n. 8, 2021.

VINÍCIUS-SILVA, R.; FERRAZ, N. P.; WERNECK, M. do V. Levantamento da fauna de abelhas (hymenoptera, apoidea) em um fragmento de mata atlântica, no município de ubá, minas gerais: Survey of the bee fauna (hymenoptera, apoidea) in an atlantic forest fragment at ubá municipality, minas gerais state, southeastern brazil. **Latin American Journal of Development**, v. 3, n. 5, p. 3203–3216, 2021.

WANG, F. **Space and Space-Time Modeling of Directional Data**. Tese (PhD Dissertation) — Duke University, 2013.