UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I
CENTRO CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE BACHARELADO EM ESTATÍSTICA

RODOLFO PEREIRA FRANKLIN

# DIREÇÃO DOS VENTOS EM CAMPINA GRANDE-PB: MODELAGEM E ANÁLISE COM DADOS DIRECIONAIS

CAMPINA GRANDE - PB
2024

**RODOLFO PEREIRA FRANKLIN**

# DIREÇÃO DOS VENTOS EM CAMPINA GRANDE-PB: MODELAGEM E ANÁLISE COM DADOS DIRECIONAIS

Trabalho de Conclusão de Curso (Monografia) apresentado ao Departamento do Curso de Estatística da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Estatística.

**Área de concentração**: Estatística Direcional

**Orientadora**: Profa. Dra. Divanilda Maia Esteves

**CAMPINA GRANDE**

**2024**

F831d Franklin, Rodolfo Pereira.

Direção dos ventos em Campina Grande - PB modelagem e análise com dados direcionais [manuscrito] / Rodolfo Pereira Franklin. - 2024.

50 f. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2024.

"Orientação : Prof. Dra. Divanilda Maia Esteves, Departamento de Estatística - CCT".

1. Estatística Direcional. 2. Direção dos Ventos. 3. Modelos de Regressão. I. Título

21. ed. CDD 551.518

Elaborada por Lêda Cristina Diniz Andrade - CRB - 15/1032 BC

RODOLFO PEREIRA FRANKLIN

DIREÇÃO DOS VENTOS EM CAMPINA GRANDE-PB: MODELAGEM E ANÁLISE COM DADOS DIRECIONAIS

Trabalho de Conclusão de Curso (Monografia) apresentado ao Departamento do Curso de Estatística da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Estatística.

Trabalho aprovado em 21 de Novembro de 2024.

**BANCA EXAMINADORA**

DMEsteves

Profa. Dra. Divanilda Maia Esteves
(Orientadora)
Universidade Estadual da Paraíba (UEPB)

MKBSilva

Profa. Dra Michelli Karinne Barros da Silva
Universidade Federal de Campina Grande
(UFCG))

Débora de Sousa Cordeiro

Profa. Mra. Débora de Sousa Cordeiro
Universidade Estadual da Paraíba (UEPB)

Aos que me serviram de suporte e exemplo nesta jornada, DEDICO.

## AGRADECIMENTOS

Agradeço, primeiramente, a Deus, pela força e sabedoria em todos os momentos desta jornada. A minha família, em especial à minha mãe e ao meu pai, pelo amor, apoio e por estarem sempre ao meu lado. A Maria Eduarda, por seu carinho e apoio constantes, fundamentais para que eu me mantivesse firme nesse percurso. A minha orientadora Diana, que com paciência e conhecimento, guiou cada etapa deste trabalho, e a todos os meus amigos que fizeram esta caminhada mais leve e significativa. Em especial, ao Gabriel, por todo o apoio e ajuda ao longo do curso; ao Antônio, que sempre me incentivou e esteve disposto a me ajudar; ao José Davi, que me auxiliou financeiramente em momentos desafiadores, permitindo que eu pudesse participar de eventos acadêmicos importantes; ao Artur, pela presença constante e apoio nos momentos difíceis; ao Vitor, por toda a força e companheirismo durante esta jornada; ao Amauri, colega de turma sempre disposto a colaborar; ao Joaquim, pela amizade sincera e por sempre me incentivar; ao Lucas, colega de turma, pela amizade, e aos professores do departamento que foram essenciais nessa trajetória. Um agradecimento especial ao professor Tiago, que em momentos cruciais estendeu sua mão e, sem o qual, talvez eu não estivesse aqui hoje; e ao professor Ricardo, por seus conselhos valiosos e pelas conversas que tanto me ajudaram. A todos que contribuíram de alguma forma, minha eterna gratidão.

“Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom” (Stoll, 2006).

**RESUMO**

Este trabalho investiga a aplicação da estatística direcional na análise dos padrões dos ventos em Campina Grande, Paraíba, com o intuito de fornecer informações úteis para áreas como o planejamento de energia eólica e o urbanismo sustentável. A estatística direcional é crucial para dados cíclicos ou angulares, como os de direção de vento, pois trata de forma adequada a natureza circular destes dados, onde ângulos de 0° e 360° representam o mesmo ponto. Os dados analisados foram coletados pela estação meteorológica do Instituto Nacional de Meteorologia (INMET) durante o primeiro semestre de 2022, totalizando 4.344 observações. Este período foi escolhido por garantir dados completos e confiáveis, permitindo uma análise robusta e, futuramente, a comparação com dados de outros períodos para identificação de padrões sazonais. O trabalho aplica métodos de estatística circular, começando com uma análise descritiva que calcula medidas de posição (direção média e mediana) e dispersão (variância e desvio padrão circular). A direção média foi calculada usando vetores unitários correspondentes aos ângulos, respeitando a estrutura cíclica dos dados. Em seguida, foram aplicados testes de uniformidade, como os de Rayleigh e Watson, para verificar se os ventos seguem uma distribuição uniforme e se há predominância de certas direções. Também foram utilizados modelos circulares, como a distribuição de von Mises, que é análoga à distribuição normal em estatísticas lineares, e a distribuição cardioide, adequada para dados com baixa concentração. Estes modelos ajudaram a entender a distribuição dos ventos em torno de uma direção média, com a distribuição de von Mises sendo especialmente útil para dados com alta concentração direcional. A análise revelou uma predominância de ventos do leste, com concentração entre 135° e 175°, sugerindo a influência de fatores geográficos e climáticos locais. Os resultados obtidos podem apoiar o planejamento urbano, auxiliando na orientação de construções para maximizar a ventilação natural, na agricultura, para a otimização de práticas de irrigação, e na energia eólica, contribuindo para a localização estratégica de turbinas. Este estudo oferece uma contribuição relevante para a estatística, a meteorologia e o desenvolvimento sustentável da região.

**Palavras-chave:** Estatística Direcional. Direção dos Ventos. Modelos de Regressão.

## ABSTRACT

This study investigates the application of directional statistics in analyzing wind patterns in Campina Grande, Paraíba, aiming to provide useful information for areas such as wind energy planning and sustainable urban development. Directional statistics are essential for cyclic or angular data, like wind direction data, as they appropriately address the circular nature of these data, where angles of 0° and 360° represent the same point. The analyzed data were collected by the meteorological station of the National Institute of Meteorology (INMET) during the first half of 2022, totaling 4,344 observations. This period was chosen to ensure complete and reliable data, allowing for robust analysis and, in the future, comparison with data from other periods to identify seasonal patterns. The study applies circular statistical methods, beginning with a descriptive analysis that calculates position measures (mean and median direction) and dispersion (circular variance and standard deviation). The mean direction was calculated using unit vectors corresponding to the angles, respecting the data's circular structure. Then, uniformity tests, such as Rayleigh and Watson tests, were applied to verify if the wind directions follow a uniform distribution and if certain directions predominate. Circular models were also used, including the von Mises distribution, analogous to the normal distribution in linear statistics, and the cardioid distribution, suitable for data with low concentration. These models helped understand how wind directions distribute around a mean direction, with the von Mises distribution being particularly useful for data with high directional concentration. The analysis revealed a predominance of eastward winds, with concentration between 135° and 175°, suggesting the influence of local geographic and climatic factors. The results can support urban planning by guiding building orientation to maximize natural ventilation, assist agriculture by optimizing irrigation practices, and aid wind energy development by informing the strategic placement of turbines. This study makes a valuable contribution to statistics, meteorology, and the sustainable development of the region.



**Keywords:** Directional Statistics. Wind Direction. Regression Models.

SUMÁRIO

# 1 INTRODUÇÃO

A estatística direcional é um ramo fascinante e complexo da estatística, que lida com dados cujo domínio não é linear, mas sim direcional ou angular. "Eles surgem comumente em Biologia, Geografia, Geologia, Geofísica, Medicina, Meteorologia e Oceanografia, e em muitas outras áreas" (Fisher, 1995, p. XV, Tradução nossa). Diferentemente dos dados tradicionais, que podem ser representados ao longo de uma linha reta (ou em um plano cartesiano), os dados direcionais são mais adequadamente representados em um círculo ou esfera, tornando suas análises estatísticas únicas e desafiadoras.

Uma das principais características dos dados direcionais é que eles são cíclicos. Por exemplo, um ângulo de 0° é o mesmo que um ângulo de 360°. Dada a forma como os dados são melhor representados, "são necessários métodos direcionais especiais que levem em conta a estrutura desses espaços amostrais" (Mardia; Jupp, 2000, P. XIX, Tradução nossa). Apesar dos dados circulares serem medidos em graus, às vezes é útil medi-los em radianos, no qual suas medidas angulares podem ser convertidas de graus para radianos multiplicando por $\pi/180$.

Um dos primeiros passos no estudo da estatística direcional pode ser rastreado até o trabalho de Mardia (1972), que estabelece as bases teóricas para a análise de dados direcionais. Ele introduz conceitos fundamentais como a distribuição de von Mises, que possui uma importância análoga à distribuição normal em estatísticas lineares, mas definida em um espaço circular.

A importância da estatística direcional é evidenciada em sua aplicação prática:

> Em diversos campos científicos, as medidas são direções. Por exemplo, um biólogo pode estar medindo a direção do voo de um pássaro ou a orientação de um animal, enquanto um geólogo pode estar interessado na direção do pólo magnético da Terra. Tais direções podem ser em duas dimensões como nos dois primeiros exemplos ou em três dimensões como o último (Jammalamadaka; Sengupta, 2001, p. 01, Tradução nossa).

Um exemplo clássico de tipos de dados na área da geologia é o estudo da orientação de cristais em seções perpendiculares à xistosidade, que mostra a melhora no alinhamento dos cristais causada por deformação progressiva (Borradaile, 2003). Outro exemplo é o uso de dados de petrofábrica, que dizem respeito à orientação dos eixos cristalográficos ou das dimensões longas dos grãos, onde os eixos não possuem sentido de direção, sendo qualquer um dos dois azimutes possíveis igualmente válidos (Borradaile, 2003).

A análise de dados direcionais não se limita apenas ao plano bidimensional. A estatística direcional tridimensional é igualmente importante, como demonstrado por Mardia e Jupp (2000), em seu trabalho sobre dados esféricos. Essa extensão é crucial em campos como a astrofísica, no qual os objetos são posicionados e movem-se em um espaço tridimensional.

Apesar de sua aplicabilidade, a estatística direcional ainda é um campo relativamente menos explorado e compreendido quando comparado a outras áreas da Estatística, lhe transformando-a em uma grande oportunidade para pesquisas e uma área com muito potencial de desenvolvimento e inovação.

À medida que se avança, torna-se cada vez mais evidente que a estatística direcional desempenhará um papel crucial em muitos campos científicos. Sua capacidade de extrair resultados significativos de dados complexos e orientados é inestimável. Assim, o estudo e a compreensão dessa área pode ser um diferencial para cientistas e pesquisadores que buscam fazer pleno uso dos dados à sua disposição.

A energia eólica, uma das principais formas de energia sustentável, desempenha um papel crucial na transição global para um futuro mais sustentável e menos dependente de combustíveis fósseis. Esta transição não é apenas uma questão de inovação tecnológica; ela é impulsionada por uma compreensão profunda e contínua dos padrões de vento, sua variabilidade e previsibilidade.

O vento, como uma fonte de energia, oferece vantagens significativas: é abundante, renovável e não emite gases de efeito estufa durante a geração de eletricidade. No entanto, sua natureza intermitente e imprevisível apresenta desafios únicos. Um estudo detalhado dos ventos abrangendo sua velocidade, direção e padrões sazonais é fundamental para otimizar o aproveitamento da energia eólica. Compreender esses padrões não apenas facilita a localização estratégica de turbinas eólicas, mas também melhora a eficiência e a confiabilidade das previsões de geração de energia.

No cenário atual, marcado pelo avanço das tecnologias de *big data* e aprendizado de máquina, o estudo dos ventos também se beneficia de análises de dados mais sofisticadas. Estas tecnologias permitem não apenas uma modelagem mais precisa dos padrões de vento, mas também uma melhor integração da energia eólica nos sistemas de energia, otimizando tanto a geração quanto a distribuição de energia.

Deste modo, o objetivo geral deste estudo é realizar uma análise e modelagem dos dados de direção dos ventos na cidade de Campina Grande, Paraíba, coletados no observatório A313 durante o ano de 2022, tendo como objetivos específicos:

a) visualizar o comportamento dos dados via estatística descritiva circular;

b) realizar testes de uniformidade a fim de verificar se os dados da direção dos ventos de fato seguem uma distribuição uniforme; e

c) realizar a modelagem via regressão circular-linear dos dados das direções dos ventos.

Esta pesquisa visa não apenas compreender os padrões predominantes de vento na região, mas também desenvolver modelos preditivos que possam fornecer resultados valiosos para diversas aplicações práticas. A cidade de Campina Grande, com suas características climáticas e geográficas únicas, apresenta um cenário ideal para estudos eólicos. A análise desses dados, utilizando métodos da estatística direcional, permitirá uma compreensão mais profunda da variação espacial e temporal dos ventos. O intuito é gerar informações cruciais que podem ser aplicadas na avaliação do potencial para geração de energia eólica. Este estudo não apenas contribuirá para o campo acadêmico da Estatística e Meteorologia, mas também pode fornecer dados valiosos para impulsionar o uso de energia renovável e sustentável na região.

# 2 ESTATÍSTICAS DESCRITIVAS CIRCULARES

Como mencionado anteriormente, devido à natureza dos dados circulares, a análise estatística tradicional pode não capturar plenamente suas características. Portanto, neste capítulo, serão apresentadas técnicas estatísticas direcionais que são mais apropriadas para esse tipo de dado e podem proporcionar resultados mais precisos e informativos.

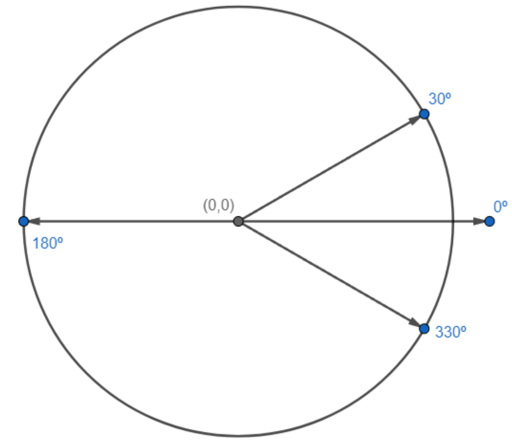
"Uma observação circular pode ser considerada como um ponto em um círculo de raio unitário, ou um vetor unitário (ou seja, uma direção) no plano. Uma vez que uma direção inicial e uma orientação do círculo tenham sido escolhidas, cada observação circular pode ser especificada pelo ângulo da direção inicial até o ponto no círculo correspondente à observação" (Mardia; Jupp, 2000, p. 01, Tradução nossa). Quando se é falado sobre ângulos na estatística direcional destacamos que normalmente o interesse para o estudo é sobre a direção e não a magnitude do vetor.

## 2.1 Medidas de Posição

### 2.1.1 Direção Média

Quando são estudados conjuntos de dados circulares unimodais, quase sempre se inicia realizando uma análise exploratória dos dados, incluindo medidas como a direção média. De acordo com os conhecimentos que se tem sobre a Estatística, seria usual calcular a média aritmética dos dados. Supondo que haja, para estudos, dados de duas medidas sendo elas $30°$ e $330°$, que representam a direção do vento na cidade de Campina Grande, Paraíba, em um determinado dia e hora, coletados na estação de medição A313, e desejando calcular a direção média dos mesmos, seria natural calcular que sua direção média é de $180°$. Pode-se visualizar melhor essas informações na Figura 1, onde fica evidenciado que o uso desta forma de calcular a direção média é equivocado, pois aponta para uma direção inadequada.

Figura 1 – Representação gráfica da média aritmética em dados circulares.




Fonte: Elaborado pelo autor, 2023.

Interpretações inadequadas podem surgir ao utilizar diversas métricas estatísticas, como média convencional, mediana, moda, variância e desvio padrão. Para calcular corretamente a direção média de dados circulares, é essencial tratar as observações (ângulos) como vetores unitários. Dessa forma, a direção média é determinada pela direção do vetor soma, resultante da combinação desses vetores unitários. Esta abordagem leva em consideração a natureza cíclica dos dados, fornecendo um resultado mais preciso e representativo.

Assim, sejam os vetores unitários $x_1, \ldots, x_n$ com ângulos correspondentes $\theta_i$ com $i = 1, \ldots, n$. A direção média $\bar{\theta}$ de $\theta_1, \ldots, \theta_n$ é a direção da resultante $x_1 + \ldots + x_n$ de $x_1, \ldots, x_n$. É também a direção do centro de massa $\bar{x}$ de $x_1, \ldots, x_n$. Como as coordenadas cartesianas de $x_j$ são $(cos\left(\theta_j\right), sen\left(\theta_j\right))$ para $j = 1, \ldots, n$, as coordenadas do centro de massa são $\left(\bar{C}, \bar{S}\right)$, no qual

$$\bar{C} = \frac{1}{n}\sum_{j=1}^{n} cos\theta_j, \quad \bar{S} = \frac{1}{n}\sum_{j=1}^{n} sen\theta_j. \tag{2.1}$$

Portanto, $\theta$ é a solução das equações

$$\bar{C} = \bar{R}cos\bar{\theta}, \quad \bar{S} = \bar{R}sen\bar{\theta}, \quad \bar{R} > 0 \tag{2.2}$$

em que, o comprimento médio resultante $\bar{R}$ é dado por

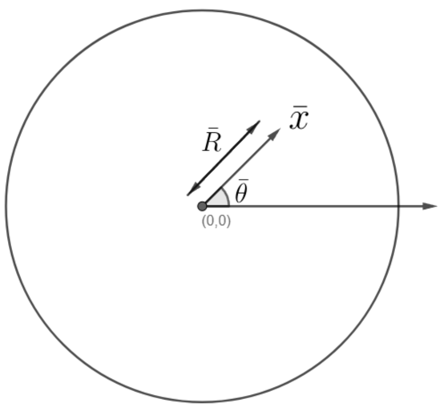
$$\bar{R} = \left(\bar{C}^2 + \bar{S}^2\right)^{1/2}. \tag{2.3}$$

Note que $\bar{\theta}$ não é definido quando $\bar{R} = 0$. Quando $\bar{R} > 0$, $\bar{\theta}$ é dado explicitamente por

$$\bar{\theta} = \begin{cases} \tan^{-1}\left(\frac{\bar{S}}{\bar{C}}\right) & \text{se } \bar{C} \geq 0, \\ \tan^{-1}\left(\frac{\bar{S}}{\bar{C}}\right) + \pi & \text{se } \bar{C} < 0, \end{cases} \tag{2.4}$$

em que a função tangente inversa "$\tan^{-1}$" (ou 'arctan') assume valores em $[-\pi/2, \pi/2]$. Observa-se que quando tratamos de estatística circular $\bar{\theta}$ não significa $(\theta_1 + \ldots + \theta_n)/n$ (o que não é bem definido, pois depende de onde o círculo é cortado). Um exemplo da representação da média direcional $\bar{\theta}$ e do comprimento resultante $\bar{R}$ é apresentado na Figura 2.

Figura 2 – Representação gráfica da média direcional e comprimento resultante.




Fonte: Elaborado pelo autor, 2023.

Também pode-se aplicar efeito de rotações na direção média da amostra. Supondo que uma nova direção inicial seja escolhida, fazendo um ângulo $\alpha$ com a direção inicial original. Então os pontos de dados correspondem a ângulos

$$\theta_j' = \theta_j - \alpha, \quad j = 1, \ldots, n, \tag{2.5}$$

neste novo sistema de coordenadas. Coloque

$$\bar{C}' = \frac{1}{n}\sum_{j=1}^{n} cos\theta_j', \quad \bar{S}' = \frac{1}{n}\sum_{j=1}^{n} sen\theta_j'. \tag{2.6}$$

Então

$$\bar{C}' = \bar{R}cos\left(\bar{\theta} - \alpha\right), \quad \bar{S}' = \bar{R}sen\left(\bar{\theta} - \alpha\right). \tag{2.7}$$

Se as coordenadas polares de $\left(\bar{C}', \bar{S}'\right)$ são $\left(\bar{R}', \bar{\theta}'\right)$, então

$$\bar{C}' = \bar{R}'cos\bar{\theta}', \quad \bar{S}' = \bar{R}'sen\bar{\theta}'. \tag{2.8}$$

A combinação de 2.7 e 2.8 dá

$$\bar{\theta}' = \bar{\theta} - \alpha, \quad \bar{R}' = \bar{R}. \tag{2.9}$$

Assim, a direção média de $\theta_1 - \alpha, \ldots, \theta_n - \alpha$ é $\bar{\theta} - \alpha$, ou seja, a direção média da amostra é equivariante sob rotação.

### 2.1.2 Direção Mediana

Quando se trata de realizar uma estimação robusta é útil ter uma versão da mediana amostral que trate adequadamente os dados circulares. Uma direção mediana da amostra $\bar{\theta}$ dos ângulos $\theta_1, \ldots, \theta_n$ é qualquer ângulo $\phi$ que:

(i) Metade dos pontos de dados estejam no arco $[\phi, \phi + \pi)$;

(ii) A maioria dos pontos de dados seja mais próximo de $\phi$ do que de $\phi + \pi$.

Quando o tamanho da amostra $n$ é ímpar, a mediana da amostra é um dos pontos de dados. Quando $n$ é par, é coveniente tomar a mediana da amostra como o ponto médio de dois pontos de dados adjacentes apropriados.

## 2.2 Medidas de Concentração e Dispersão

### 2.2.1 Variância Circular

Como já foi visto anteriormente em 2.3, o comprimento médio resultante $\bar{R}$ que foi introduzido como o comprimento do vetor do centro de massa $\bar{x}$ e é dado por

$$\bar{R} = (\bar{C}^2 + \bar{S}^2)^{1/2}.$$

Como $x_1, \ldots, x_n$ são vetores unitários,

$$0 \leq \bar{R} \leq 1. \tag{2.10}$$

Observa-se que se as direções dos ângulos $\theta_1, \ldots, \theta_n$ estiverem fortemente agrupadas, então $\bar{R}$ será quase 1, por outro lado, se as direções dos ângulos estiverem amplamente dispersas o $\bar{R}$ será quase 0. Tendo isto em mente evidencia-se que o $\bar{R}$ é uma medida de concentração dos dados. Observe que qualquer conjunto de dados da forma $\theta_1, \ldots, \theta_n$, $\theta_1 + \pi, \ldots, \theta_n + \pi$ tem $\bar{R} = 0$. É importante destacar que $\bar{R} \approx 0$ não implica necessariamente que as direções estejam distribuídas quase uniformemente ao redor do círculo. Isso ocorre porque o comprimento resultante médio próximo de zero também pode resultar de direções que estão uniformemente distribuídas em duas antípodas, ou opostas, no círculo. Por exemplo, se você tem um conjunto de dados em que cada direção $\theta_i$ é acompanhada por outra direção que é $\theta_i + \pi$ (ou seja, a direção oposta no círculo), o vetor soma dessas direções será muito pequeno ou próximo de zero, porque elas se cancelam mutuamente. Isso não significa que as direções estão uniformemente espalhadas ao redor do círculo, mas sim que existem duas direções predominantes opostas que resultam em um comprimento resultante médio pequeno.

O comprimento resultante $R$ é o comprimento do vetor resultante $x_1 + \ldots + x_n$, onde desta forma:

$$R = n\bar{R}. \tag{2.11}$$

Para a maior parte das análises descritivas e inferenciais, a métrica de comprimento médio resultante $\bar{R}$ é mais significativa do que outras medidas de dispersão. Contudo, quando se é comparado com dados lineares, pode ser vantajoso levar em conta medidas de dispersão específicas para dados circulares. Entre essas medidas, a mais básica é a variância circular da amostra, que é definida como

$$V = 1 - \bar{R}. \tag{2.12}$$

Segue de 2.10 que

$$0 \leq V \leq 1. \tag{2.13}$$

### 2.2.2 Desvio Padrão Circular

Em certas ocasiões, pode ser proveitoso dispor de um paralelo ao desvio padrão dos dados lineares para dados circulares. Uma estratégia para alcançar essa estatística resumida envolve a conversão da variância amostral $V$. Então tem-se que o desvio padrão circular é dado por:

$$v = \{-2log\,(1-V)\}^{1/2} = \left\{-2log\left(\bar{R}\right)\right\}^{1/2}. \tag{2.14}$$

### 2.2.3 Dispersão Circular

A dispersão circular amostral $\hat{\delta}$ é

$$\hat{\delta} = \frac{1-\bar{R}_2}{2\bar{R}^2}, \tag{2.15}$$

em que $\bar{R}_2$ denota o comprimento médio resultante dos ângulos $2\theta_1, \ldots, 2\theta_n$.

Entre as medidas de dispersão mencionadas, a variância circular e o desvio padrão circular são as mais comumente utilizadas em análises de dados circulares. A variância é frequentemente preferida devido à sua simplicidade e facilidade de cálculo, enquanto o desvio padrão é escolhido por sua capacidade de fornecer uma interpretação mais intuitiva, especialmente quando se está comparando dados circulares a dados lineares. A escolha da medida de dispersão mais vantajosa pode depender do contexto específico da análise, da distribuição dos dados e dos objetivos do estudo. Portanto, é aconselhável considerar as características dos dados e a audiência ao comunicar os resultados.

# 3 MODELOS CIRCULARES

Aqui serão fornecidos alguns conceitos teóricos básicos e alguns modelos para dados circulares.

## 3.1 A Função de Distribuição

Uma abordagem para definir uma distribuição sobre o círculo unitário é através da sua função de distribuição. Partindo de uma direção inicial e uma orientação estabelecida para o círculo unitário, pode-se tratar a distribuição como a de uma variável angular aleatória $\theta$. A função de distribuição $F$ é, então, especificada como uma função definida para todo $\mathbb{R}$, conforme descrito a seguir.

$$\begin{aligned} F(x) &= \Pr(0 \leq \Theta \leq x), \qquad 0 \leq x \leq 2\pi, \\ F(x+2\pi) - F(x) &= 1, \qquad -\infty < x < \infty. \end{aligned} \tag{3.1}$$

A Equação 3.1 simplesmente declara que qualquer segmento de comprimento $2\pi$ no círculo unitário possui probabilidade 1, considerando que tal segmento representa a circunferência completa do círculo.

Para $\alpha \leq \beta \leq \alpha + 2\pi$,

$$Pr\left(\alpha < \Theta \leq \beta\right) = F\left(\beta\right) - F\left(\alpha\right) = \int_{\alpha}^{\beta} dF\left(x\right), \tag{3.2}$$

no qual a integral mencionada é uma integral de Lebesgue-Stieltjes. A função de distribuição $F$ é uma função contínua à direita. Diferentemente das funções de distribuição na linha real,

$$\lim_{x\to\infty} F\left(x\right) = \infty, \qquad \lim_{x\to-\infty} F\left(x\right) = -\infty$$

e, por definição,

$$F\left(0\right) = 0, \qquad F\left(2\pi\right) = 1.$$

Caso a função de distribuição $F$ seja absolutamente contínua, então ela possui uma função de densidade de probabilidade $f$ correspondente, de tal forma que

$$\int_{\alpha}^{\beta} f(\theta)\, d\theta = F(\beta) - F(\alpha), \qquad -\infty < \alpha \leq \beta < \infty.$$

Uma função $f$ é a função densidade de probabilidade de uma distribuição absolutamente contínua se e somente se

(i) $f(\theta) \geq 0$ em quase todos os lugares de $(-\infty, \infty)$;

(ii) $f(\theta + 2\pi) = f(\theta)$ em quase todos os lugares de $(-\infty, \infty)$;

(iii) $\int_0^{2\pi} f(\theta)\, d\theta = 1$.

## 3.2 Distribuições em Rede

Considere uma distribuição discreta com

$$P_r\left(\Theta = v + \frac{2\pi r}{m}\right) = P_r, \qquad r = 0, 1, \ldots, m-1, \tag{3.3}$$

e

$$P_r \geq 0, \qquad \sum_{r=0}^{m-1} Pr = 1.$$

Os pontos $v+2\pi r/m$ representam os vértices de um polígono regular com $m$ lados, inscrito no círculo unitário. No caso de todos os pesos serem iguais, então

$$P_r = \frac{1}{m}. \tag{3.4}$$

Essa distribuição é chamada de distribuição uniforme discreta em $m$ pontos.

## 3.3 Distribuição Uniforme

A distribuição mais elementar em um círculo é a uniforme, frequentemente adotada como o modelo de referência. Essa distribuição se destaca por ser a única no círculo que mantém a invariância tanto em rotações quanto em reflexões. Sua função de densidade de probabilidade é caracterizada por

$$f(\theta) = \frac{1}{2\pi}. \tag{3.5}$$

Assim, para $\alpha \leq \beta \leq \alpha + 2\pi$

$$Pr\left(\alpha < \theta \leq \beta\right) = \frac{\beta - \alpha}{2\pi},$$

ou seja, a probabilidade é proporcional ao comprimento do arco.

### 3.3.1 Direção Média da Distribuição Uniforme

Na estatística direcional, a direção média de uma distribuição uniforme é um conceito interessante, dado que a distribuição é isotrópica, ou seja, não favorece nenhuma direção específica. Para uma distribuição uniforme no círculo, a direção média não é bem definida da mesma forma que em distribuições unimodais.

Dado que a função de densidade de probabilidade é constante ao longo do círculo, a direção média, quando calculada, resulta em um valor que não reflete uma preferência direcional. Isso se deve ao fato de que as contribuições de ângulos opostos se cancelam. Em termos matemáticos, a direção média pode ser expressa como:

$$\bar{\theta} = \arg\left(\sum_{i=1}^{n} e^{i\theta_i}\right), \tag{3.6}$$

no qual $\theta_i$ são os ângulos de dados. Para uma amostra proveniente de uma distribuição uniforme, o vetor resultante se aproxima de zero, levando a uma direção média que não é representativa.

Essa propriedade destaca a natureza distintiva da distribuição uniforme em relação a outras distribuições circulares, em que a direção média pode fornecer informações valiosas sobre a tendência dos dados.

## 3.4 Distribuição Cardióide

A modificação da densidade uniforme por uma função cosseno produz a distribuição cardióide $C\left(\mu, \rho\right)$, cuja função de densidade de probabilidade é definida pela seguinte expressão:

$$f\left(\theta\right) = \frac{1}{2\pi}\left\{1 + 2\rho\cos\left(\theta - \mu\right)\right\}, \qquad |\rho| < \frac{1}{2}. \tag{3.7}$$

O nome provém do fato de que a curva, dada em coordenadas polares por $r = f\left(\theta\right)$ com $f$ assim como em 3.7, é uma curva cardióide. A distribuição foi introduzida por Jeffreys (1948).

O comprimento médio resultante da distribuição cardióide $C(\mu, \rho)$ é dado por $\rho$ e, se

$\rho > 0$, a direção média corresponde a $\mu$. Caracteriza-se por ser uma distribuição simétrica e unimodal, com a moda localizada em $\mu$ condicionalmente a $\rho > 0$. Quando $\rho = 0$, a distribuição cardióide se reduz à distribuição uniforme. Valores pequenos de $\rho$ indicam que a $C(\mu, \rho)$ representa um ligeiro afastamento da uniformidade. As distribuições cardióides são comumente utilizadas como aproximações para baixa concentração na distribuição de von Mises.

### 3.5 Distribuições Normais Projetadas

As distribuições no círculo podem ser obtidas pela projeção radial das distribuições no plano. Dado que $x$ é um vetor bidimensional aleatório tal que $P_r(x = 0) = 0$. Então $||\mathbf{x}||^{-1}\mathbf{x}$ é um ponto aleatório no círculo unitário. Uma instância importante é aquela em que $\mathbf{x}$ tem uma distribuição normal bivariada $N_2(\mu, \mathbf{\Sigma})$, caso em que $||\mathbf{x}||^{-1}\mathbf{x}$ é dito ter uma distribuição $PN_2(\mu, \mathbf{\Sigma})$ normal projetada (ou gaussiana angular ou normal de deslocamento).

É possível mostrar que a função de densidade de probabilidade da distribuição normal projetada $PN_2(\mu, \mathbf{\Sigma})$ é

$$p(\theta; \mu, \mathbf{\Sigma}) = \frac{\phi(\theta; \mu, \mathbf{\Sigma}) + |\mathbf{\Sigma}|^{-1/2} D(\theta)\, \mathbf{\Phi}(D(\theta))\, \phi\left(|\mathbf{\Sigma}|^{-1/2}\left(\mathbf{x}^\top \mathbf{\Sigma}^{-1}\mathbf{x}\right)^{-1/2} \mu \wedge \mathbf{x}\right)}{\mathbf{x}^\top \mathbf{\Sigma}^{-1} \mathbf{x}}, \tag{3.8}$$

em que $\phi(\cdot; \mu, \mathbf{\Sigma})$ denota a função de densidade de probabilidade de $N_2(0, \mathbf{\Sigma})$, $\phi$ e $\Phi$ denotam a função de densidade de probabilidade e a função de densidade cumulativa de $N(0,1)$, $x = (cos(\theta), sen(\theta))^\top$,

$$D(\theta) = \frac{\mu^\top \mathbf{\Sigma}^{-1} \mathbf{x}}{\mathbf{x}^\top \mathbf{\Sigma}^{-1} \mathbf{x}^{1/2}}$$

e $\mu \wedge x = 1 sen(\theta) - \mu_2 \cos(\theta)$ com $\mu = (\mu_1, \mu_2)^\top$. Em particular,

$$p(\theta; (\mu, 0), \mathbf{I}_2) = \frac{1}{\sqrt{2\pi}} \phi(\mu) + \mu \cos(\theta) \phi(\mu sen(\theta))\, \mathbf{\Phi}(\mu \cos(\theta)), \tag{3.9}$$

em que $\mathbf{I}_2$ denota a matriz identidade $2 \times 2$. A distribuição $PN_2(\mu, \mathbf{\Sigma})$ se reduz à distribuição uniforme se e somente se $\mu = 0$ e $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_2$. As distribuições normais projetadas podem ser bimodais e/ou assimétricas.

As distribuições normais projetadas com $\mu = 0$ são chamadas de distribuições gaussianas centrais angulares. A função de densidade de probabilidade da distribuição gaussiana central angular $PN_2\left(0, \mathbf{\Sigma}\right)$ é

$$p\left(\theta; \mathbf{\Sigma}\right) = \frac{\left(1 - b^2\right)^{1/2}}{2\pi\left(1 - b\cos(2\left(\theta - \mu\right))\right)}, \tag{3.10}$$

em que

$$b = \frac{2\left(tr\left(\mathbf{\Sigma}\right) - 2|\mathbf{\Sigma}|^{1/2}\left(tr\left(\mathbf{\Sigma}\right) + 2|\mathbf{\Sigma}|^{1/2}\right)^{3/2}\right)}{\left(2tr\mathbf{\Sigma}\right)^2 + 4|\mathbf{\Sigma}|}, \qquad \tan(\mu) = \frac{2\sigma_{12}}{\sigma_{11} - \sigma_{22}}$$

com

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

Observe que $p\left(\theta; c\mathbf{\Sigma}\right) = p\left(\theta; \mathbf{\Sigma}\right)$ para qualquer $c$ diferente de zero, então podemos assumir sem perda de generalidade que $|\mathbf{\Sigma}| = 1$. Como $p\left(\theta; \mathbf{\Sigma}\right) = p\left(\theta + \pi; \mathbf{\Sigma}\right)$, as distribuições gaussianas centrais angulares fornecem modelos úteis para dados axiais.

### 3.6 Distribuição de Von Mises

Do ponto de vista da inferência estatística, talvez a distribuição mais útil no círculo seja a distribuição de Von Mises, foi introduzida por Richard von Mises em 1918 em seu trabalho seminal sobre a teoria da probabilidade (Von Mises, 1918). No artigo de (Fisher; Lee, 1992), a distribuição foi empregada em modelos de regressão para respostas angulares, onde o objetivo era prever direções com base em variáveis explicativas contínuas, destacando a adequação da distribuição de Von Mises em situações de concentração em torno de uma direção média, típica de dados circulares (Fisher; Lee, 1992). Além disso, (Mardia, 1972) explorou a distribuição de Von Mises em seu livro clássico sobre dados direcionais, no qual examinou suas propriedades e aplicou a distribuição em vários contextos, como a biologia, para modelar movimentos e orientações de animais, e em geologia, para a análise de direções de falhas e dobras tectônicas (Mardia, 1972). A capacidade da distribuição de Von Mises de modelar dados circulares com concentração em torno de uma direção específica a torna uma escolha natural em análises em que os dados apresentam um padrão cíclico, como nas áreas de meteorologia (direções do vento), engenharia (orientação de objetos), e psicologia (análise de respostas circulares).

### 3.6.1 Definição

A distribuição de Von Mises $M(\mu, k)$ é caracterizada por uma função de densidade de probabilidade, que é expressa como:

$$g(\theta; \mu, k) = \frac{1}{2\pi I_0(k)} e^{k\cos(\theta-\mu)}, \quad \theta \in [0, 2\pi). \tag{3.11}$$

Nesta expressão, $I_0$ denota a função de Bessel modificada de primeiro tipo e ordem zero. A função de Bessel modificada de primeiro tipo para uma ordem $p$ qualquer é apresentada em (Abramowitz; Stegun, 1965, p. 376). e é representada pela seguinte fórmula:

$$I_p(k) = \frac{1}{2\pi} \int_0^{2\pi} \cos(p\theta) e^{k\cos(\theta)} d\theta \tag{3.12}$$

Contudo, especificamente para o nosso estudo, consideramos a função de Bessel de ordem zero. Portanto, ao substituir $p$ por zero na fórmula acima, obtemos:

$$\begin{aligned} I_0(k) &= \frac{1}{2\pi} \int_0^{2\pi} \cos(0\theta) \cdot e^{k\cos(\theta)} d\theta \\ I_0(k) &= \frac{1}{2\pi} \int_0^{2\pi} e^{k\cos(\theta)} d\theta \end{aligned} \tag{3.13}$$

A função $I_0$ tem expansão em série de potência dada por:

$$I_0(k) = \sum_{r=0}^{\infty} \frac{1}{(r!)^2} \left(\frac{k}{2}\right)^{2r} \tag{3.14}$$

O parâmetro $\mu$ é a direção média e o parâmetro $k$ é conhecido como parâmetro de concentração. O comprimento médio resultante $\rho$ é $A(k)$, em que $A$ é dada por:

$$A(k) = I_1(k) / I_0(k). \tag{3.15}$$

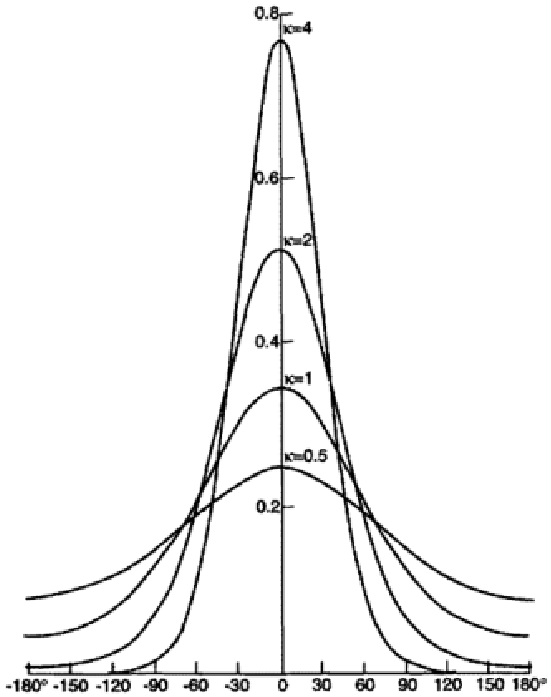
Observa-se que as distribuições $M(\mu + \pi, k)$ e $M(\mu, -k)$ são, na verdade, idênticas. Para resolver essa ambiguidade nos parâmetros $\mu$ e $k$, é comum adotar a convenção de que $k \geq 0$.

### 3.6.2 A Forma da Distribuição

A distribuição é unimodal e simétrica em torno de $\theta = \mu$. A moda está em $\theta = \mu$ e a antimoda em $\theta = \mu + \pi$. A razão da densidade na moda para a densidade na antimoda é dada por $e^{2k}$, de modo que quanto maior o valor de $k$, maior é o agrupamento em torno da moda.

A Figura 3 mostra a densidade para $\mu = 0$ e $k = 0.5, 1, 2, 4$. Para $k = 4$, mais de 99% da probabilidade encontra-se no arco $(-90°, 90°)$.

Figura 3 – Gráfico da função da distribuição de von Mises $M(0,k)$ para $k = 0.5, 1, 2e4$.




Fonte: (Mardia; Jupp, 2000)

### 3.6.3 Relação com outras Distribuições

Quando $k$ é igual a zero, a distribuição $M(\mu, k)$ corresponde à distribuição uniforme. Utilizando a aproximação $exp(x) \approx 1 + x$ para valores pequenos de $k$, podemos observar que

$$M(\mu, k) \approx C(\mu, k/2), \tag{3.16}$$

no qual $C(\mu, k/2)$ denota uma distribuição cardióide, conforme definido na seção 3.4. Assim, uma distribuição de von Mises com pequeno parâmetro de concentração pode ser aproximada pela distribuição cardióide com a mesma média e comprimento médio resultante. Como $k \to \infty$, a distribuição $M(\mu, k)$ torna-se concentrada no ponto $\theta = \mu$. Se $k$ for grande, coloque $\xi = k^{1/2}(\theta - \mu)$. Então, de 3.11, a função de densidade de probabilidade de $\psi$ é proporcional a

$$exp\left\{-k\left[1-cos\left(k^{-1/2}\xi\right)\right]\right\}. \tag{3.17}$$

Para $k$ grande,

$$1-cos\left(k^{-1/2}\xi\right)=\frac{1}{2}k^{-1}\xi^2+\mathrm{O}\left(k^{-2}\right),$$

de modo que, de 3.17, $\xi \dot{\sim} N(0,1)$. Portanto, para $k$ grande

$$\theta \dot{\sim} M(\mu,k) \Rightarrow k^{-1/2}(\theta-\mu) \dot{\sim} N(0,1), k \to \infty. \tag{3.18}$$

De maneira mais geral, qualquer distribuição de von Mises pode ser aproximada por uma distribuição normal envolta, conforme será definido mais a frente. É apropriado igualar as distribuições igualando seus primeiros momentos trigonométricos. Assim, utilizando 3.10, temos

$$M(\mu,k) \cong WN(\mu,A(k)), \quad k \to \infty. \tag{3.19}$$

Embora a aproximação 3.19 tenha sido derivada aqui como uma aproximação de primeira ordem para $k$ grande, (Kent, 1978), temos

$$f_{VM}(\theta;\mu,k)-f_{VM}(\theta;\mu,A(k))=O\left(k^{-1/2}\right), \qquad k \to \infty, \tag{3.20}$$

no qual $f_{VM}(\theta;\mu,k)$ e $f_{VM}(\theta;\mu,A(k))$ denotam as densidades da distribuição de von Mises $<(\mu,k)$. Foi verificado por (Stephens, 1963) numericamente que a aproximação 3.19 é satisfatória para valores intermediários de $k$. A pior correspondência (em termos de diferença absoluta máxima das funções de densidade de probabilidade) entre uma distribuição de von Mises e uma distribuição normal envolta aproximada ocorre para $k=1,4$, porém mesmo neste caso as duas densidades são muito próximas.

### 3.6.4 A função da distribuição

A função de distribuição de von Mises $M(0,k)$ é

$$F(\theta;0,k)=\frac{1}{I_0(k)}\int_0^\theta e^{kcos(u)}du \tag{3.21}$$

e não é particularmente fácil de avaliar numericamente. O resultado 3.19 significa que a função de distribuição de $N(\mu, k^{-1})$ fornece uma aproximação para a função de distribuição de $M(\mu, k)$. Para $k > 10$, esta aproximação é muito precisa.

Aproximações aprimoradas para as funções de distribuição das distribuições de von Mises são fornecidas por transformações de normalização aproximadas que refinam a transformação $\theta \rightarrow k^{1/2}$ utilizada em 3.17. Uma aproximação foi criada por (Upton, 1974) e posteriormente, refinada por (G. W. Hill, 1976), é que se $\theta \sim M(0, k)$ então $X \dot{\sim} N(0, 1)$, em que

$$X = y - \frac{y}{8k} - \frac{2y^3 + 7y}{128k^2} - \frac{8y^5 + 46y^3 + 177y}{3072k^3} - \ldots$$

e

$$y = 2\sqrt{k} sen\left(\frac{\theta}{2}\right).$$

Algoritmos de computador para avaliar a função de distribuição de von Mises foram fornecidos por **?** e (G. W. Hill, 1977).

# 4 TESTE DE UNIFORMIDADE

A uniformidade circular ou isotropia é a hipótese básica mais fundamental em estatísticas circulares. Se ela não pode ser rejeitada, então pode-se assumir razoavelmente que a população da qual a amostra foi retirada está distribuída de maneira uniforme ao redor do círculo. Quando a uniformidade é rejeitada, os dados suportam alguma forma de desvio da uniformidade.

## 4.1 Teste de Rayleigh

Um dos testes mais simples de uniformidade é o teste de Rayleigh. Considerando que $E\left[(cos\theta, sen\theta)^{\top}\right] = 0$ quando $\theta$ tem distribuição uniforme, é intuitivamente razoável rejeitar a uniformidade quando a média amostral vetorial $\left(\bar{C}, \bar{S}\right)$ está distante de 0, ou seja, quando $\bar{R}$ é grande. (Mardia; Jupp, 2000) explica que é útil utilizar a estatística do teste como $2n\bar{R}^2$.

Formalmente uma boa justificativa para utilizar o mesmo é que ele é o teste de escore de uniformidade dentro do modelo de von Mises 3.11. Tendo $\omega = (kcos\mu, ksen\mu)^{\top}$, teremos que $\omega$ é o parâmetro natural deste modelo exponencial e a log-verossimilhança baseada em $\theta_1, \ldots, \theta_n$ é

$$l\left(\omega; \theta_1, \ldots, \theta_n\right) = n\omega^{\top}\mathbf{\bar{x}} - nlogI_0\left(k\right), \tag{4.1}$$

em que

$$\mathbf{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n}\left(cos\theta_i, sen\theta_i\right)^{\top}$$

é a média amostral do vetor. o escore é

$$\mathbf{U} = \frac{\partial l}{\partial \omega^{\top}} = n\mathbf{\bar{x}} - nA\left(k\right)\left(cos\mu, sen\mu\right)^{\top}. \tag{4.2}$$

Então, a estatística de escore é

$$\mathbf{U}^{\top}var\left(\mathbf{U}\right)^{-1}\mathbf{U} = 2n\bar{R}^2. \tag{4.3}$$

## 4.2 Teste de Watson's $U^2$

Outro teste utilizado para avaliar a uniformidade de uma distribuição de dados direcionais é o teste de Watson, diferentemente de outros testes que medem a discrepância entre a função de distribuição empírica e a função de distribuição cumulativa uniforme através do máximo desvio, o teste de Watson utiliza o desvio quadrático médio corrigido. Essa abordagem permite uma avaliação mais equilibrada das diferenças ao longo de todo o círculo, evitando a ênfase excessiva em desvios extremos. A estatística de Watson é citada em (Mardia; Jupp, 2000) e é dada por

$$U^2 = n \int_0^{2\pi} \left\{ S_n(\theta) - \frac{\theta}{2\pi} - \mu \right\}^2 \frac{1}{2\pi} d\theta, \tag{4.4}$$

em que

$$\begin{aligned} \mu &= \int_0^{2\pi} \left\{ S_n(\theta) - \frac{\theta}{2\pi} \right\} \frac{1}{2\pi} d\theta \\ &= \frac{1}{2} - \frac{1}{n} \sum_{i=1}^{n} \frac{\theta_{(i)}}{2\pi}. \end{aligned} \tag{4.5}$$

## 4.3 Regressão Circular

Com intuito de analisar o impacto de uma variável linear em nossa variável alvo do estudo, serão utilizados utilizar de modelos de regressão circular, mais especificamente o modelo circular-linear. Para isso utilizaremos da função *lm.circular* do pacote anteriormente citado *circular* do software R que nos permite explorar o ajuste de um modelo relativamente restritivo. Se possuímos um conjunto de $n$ observações independentes, $(\theta_1, x_1), \ldots, (\theta_n, x_n)$, no qual $x_j$ denota o vetor $(x_{1j}, \ldots, x_{kj})$. Assume-se que $\theta_j$ foram extraídos de uma distribuição de von Mises com direção média $\mu_j$ e parâmetro de concentração $k$, sendo que $\mu_j$ está relacionado às variáveis explicativas $X_1, \ldots, X_k$ de acordo com a equação 4.6)

$$\mu_j = \mu + g(\gamma_1 X_1 + \ldots + \gamma_k X_k), \tag{4.6}$$

no qual $\mu$ e $\gamma_1, \ldots, \gamma_k$ são constantes que devem ser estimadas, assim como $k$. Para este modelo a função de ligação

$$g(u) = 2\tan^{-1}(u) \tag{4.7}$$

é geralmente utilizada. Para esta escolha de função de ligação, $g(0) = 0$ (então $\mu$ pode ser interpretado como a origem), e $g(u)$ assume valores entre $-\pi$ e $\pi$ conforme $u$ varia de $-\infty$ a $\infty$. A função *lm.circular* usa a estimação de máxima verossimilhança para estimar os parâmetros desconhecidos. Para mais detalhes da teoria ler (Fisher; Lee, 1992) e (Fisher, 1993), sendo que esta seção se baseia fortemente nestas duas publicações.

# 5 RESULTADOS

## 5.1 Dados Meteorológicos

Os dados utilizados neste estudo foram fornecidos pelo Instituto Nacional de Meteorologia (INMET). Especificamente, os dados compreendem registros horários da direção dos ventos coletados automaticamente pela estação meteorológica situada em Campina Grande durante o primeiro semestre de 2022. Este período foi escolhido para este estudo por várias razões fundamentadas. Primeiramente, na data da coleta de dados, o ano de 2023 ainda não tinha seus dados completamente disponibilizados. Para garantir a integridade e a completude dos dados analisados, optamos por um período anterior cuja disponibilidade de informações fosse total e confiável.

Além disso, o primeiro semestre de 2022 oferece uma base sólida para análises futuras. Ao utilizar este período, podemos comparar com os dados de 2023 quando estiverem completos, possibilitando a verificação de padrões sazonais e mudanças significativas que possam ter ocorrido. Este tipo de análise longitudinal é importante para identificar tendências consistentes ou variações relevantes ao longo do tempo.

O período de seis meses escolhido também se justifica pela quantidade significativa de dados disponíveis, com um total de 4.344 observações. Essa amplitude de dados permite uma análise estatística robusta, fornecendo resultados detalhados e confiáveis sobre os padrões observados.

Campina Grande está estrategicamente localizada no Nordeste do Brasil, uma região sujeita a complexas dinâmicas atmosféricas devido à sua topografia variada e a influências oceânicas. A estação local é dotada de equipamentos modernos e precisos, capazes de realizar medições contínuas de diversas variáveis meteorológicas, o que inclui a direção dos ventos. Esta coleta sistemática e meticulosa não apenas garante medições precisas, mas também captura nuances diurnas e sazonais nos padrões dos ventos, oferecendo um panorama detalhado das condições climáticas locais.

O entendimento profundo desses padrões é vital para o estudo, visto que a direção dos ventos na região é influenciada por uma série de fatores geográficos, como a topografia local e a proximidade ao Oceano Atlântico. Esses fatores podem alterar significativamente as características dos ventos, influenciando desde o clima até aspectos mais práticos da vida cotidiana e da economia local. As informações obtidas são extremamente valiosas para diversas áreas aplicadas. No urbanismo, por exemplo, a compreensão dos padrões de vento auxilia no planejamento e no *design* de edifícios e espaços urbanos, visando otimizar a ventilação natural e reduzir o uso de sistemas de climatização artificial. Na agricultura, o conhecimento preciso da direção dos ventos é crucial para a gestão eficaz de práticas como a aspersão de pesticidas e a irrigação, minimizando perdas e maximizando a eficiência.

Os dados da base fornecida pelo INMET contêm várias colunas, abrangendo uma gama de variáveis meteorológicas. Para a construção dos modelos de regressão, as colunas mais relevantes selecionadas foram a da Temperatura Máxima e a da Velocidade do Vento. Essa escolha se justifica pela relação direta que essas variáveis têm com os padrões de vento e sua influência nas condições climáticas locais.

Este estudo busca não apenas mapear a direção dos ventos em Campina Grande, mas também interpretar como essas direções impactam as condições locais, contribuindo para uma melhor previsão meteorológica e planejamento ambiental. A análise detalhada propiciada pelos dados coletados pelo INMET permite uma melhor compreensão das interações entre os padrões de vento e outros elementos climáticos, como precipitação e temperatura, oferecendo uma base robusta para a formulação de estratégias de adaptação às mudanças climáticas que são cientificamente embasadas e especificamente ajustadas às necessidades da região.

Além disso, o estudo ajuda a fundamentar políticas de gestão ambiental e desenvolvimento sustentável, facilitando a tomada de decisões informadas que podem levar a uma melhor qualidade de vida e a um desenvolvimento econômico mais sustentável. Ao fornecer resultados detalhados e cientificamente validados sobre os padrões climáticos locais, este trabalho não apenas enriquece o corpo acadêmico de conhecimentos meteorológicos, mas também desempenha um papel importante na orientação de políticas públicas e práticas privadas em várias esferas da sociedade.

## 5.2 Análise dos Dados

Antes de realizar qualquer análise, foi necessário realizar um pré-processamento dos dados meteorológicos. Este processo envolveu as seguintes etapas:

### 5.2.1 Configuração do Ambiente de Trabalho e Carregamento dos Pacotes

Primeiramente, foi configurado o diretório de trabalho e carregados os pacotes necessários para a análise:

```
setwd("E:/Faculdade/TCC")


require(circular)
require(cplots)
library(plotrix)
```

### 5.2.2 Importação e Verificação dos Dados

Os dados meteorológicos referentes ao período de janeiro de 2022 a junho de 2022 foram importados a partir de um arquivo CSV utilizando a função `read.csv2`. Em seguida, foi realizada uma verificação dos valores ausentes em cada coluna para garantir a integridade dos dados:

```
dados <- read.csv2("Dados Metereologicos jan 2022 a jun 2022.csv")


# Contando valores ausentes em cada coluna
na_counts <- sapply(dados, function(x) sum(is.na(x)))


# Exibindo o resultado
print(na_counts)
```

Os resultados da contagem de valores ausentes indicaram que algumas colunas continham dados faltantes, especificamente nas colunas `Temp..Max...C.`, `Temp..Min...C.`, `Umi..Min.....`, `Pto.Orvalho.Max...C.`, `Pressao.Max...hPa.`, e `Pressao.Min...hPa.`. Esses valores ausentes foram tratados conforme as necessidades da análise subsequente, garantindo que os dados utilizados estivessem completos e consistentes.

### 5.2.3 Conversão das Direções dos Ventos para Dados Circulares

Para adequar a análise às características circulares das direções dos ventos, a coluna `Dir..Vento..m.s.` foi convertida para um objeto circular utilizando o pacote `circular`.

A escolha de manter as direções em graus facilita a interpretação dos resultados e a aplicação das técnicas de estatística direcional apropriadas:

```
circ_data <- dados$Dir..Vento..m.s.
circ_data <- circular(circ_data, type = "angles", units = "degrees")
```

O pré-processamento dos dados assegurou que todas as direções dos ventos fossem adequadamente convertidas para um formato circular em graus, facilitando a utilização de métodos estatísticos específicos para dados circulares. Além disso, a verificação e tratamento dos valores ausentes garantiram a integridade e a confiabilidade dos dados utilizados nas análises subsequentes, contribuindo para a robustez dos resultados obtidos.

### 5.2.4 Distribuição Pontual dos Dados

A Figura 4 apresenta a distribuição direcional das direções do vento, representadas em um gráfico circular. O círculo indica a direção do vento em relação ao norte geográfico, que está no topo do círculo, enquanto o ponto 0 está posicionado no leste geográfico. As direções são distribuídas de forma que cada ponto no gráfico representa uma observação de direção de vento no local de estudo.

Figura 4 – Distribuição Pontual dos dados.


Fonte: Do Autor, 2024

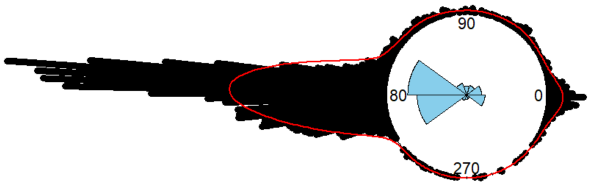
Observa-se uma maior concentração entre 135° e 175°, indicando uma predominância de ventos vindos do leste. Este padrão destaca a frequência elevada dos ventos orientais durante o período analisado, o que pode ser influenciado por características geográficas específicas ou padrões climáticos regionais que favorecem essa direção. Paralelamente, uma concentração menor de pontos em torno de 0° que sugere episódios menos frequentes de ventos vindos do oeste. Essa observação aponta para a ocorrência esporádica desses ventos, que, embora menos comuns, são relevantes para uma análise completa das dinâmicas de vento na região. O gráfico destaca a variabilidade e a predominância das direções do vento no local estudado, servindo como uma ferramenta importante para visualizar a intensidade e a direção dos ventos ao longo do tempo analisado. Esta representação direcional é relevante para entender as dinâmicas climáticas da área.

### 5.2.5 Medidas de Posição e Dispersão

Utilizando do comando *mean.circular(x)* do pacote *circular* determinamos a direção média do vento a partir das observações coletadas. Os cálculos apontaram para uma direção média de aproximadamente 3.109 radianos, equivalentes a cerca de 178.14 graus. Esta média, representada em um sistema circular onde o zero radiano é alinhado ao oeste geográfico, indica que a direção predominante dos ventos durante o período em estudo foi levemente direcionada ao leste. Este padrão não é apenas uma característica meteorológica da região, mas também um indicador climático crucial que reflete a dinâmica atmosférica local. A predominância dos ventos vindos do leste pode ser influenciada por fatores geográficos específicos da região, como a configuração do relevo e grandes corpos d'água, que modulam os padrões climáticos locais e regionais. A mediana dos dados foi de 3.106 que corresponde a aproximadamente 178 graus, esta proximidade dos valores da média e da mediana circulares indica uma distribuição simétrica das direções dos ventos ao redor do oeste. Essa consistência sugere que os ventos vieram predominantemente do leste, direcionando-se para o oeste durante o período de estudo.

É possível utilizar o diagrama da Rosa na Figura 5 para visualizar de outra forma a distribuição dos dados. Este gráfico revela uma predominância de ventos provenientes do leste, como indicado pelo setor mais largo e proeminente orientado para a esquerda do diagrama, correspondendo a 180 graus. Além disso, uma concentração menor de ventos originados do oeste é visível nos 0 e 360 graus, indicando que esses ventos são menos comuns ou mais fracos durante o período estudado.

Figura 5 – Diagrama da Rosa.




Fonte: Do Autor, 2024

O comprimento do vetor médio, também conhecido como intensidade média ou módulo do vetor médio, é uma medida crucial em estatística direcional, pois indica a concentração dos dados de direção do vento em torno da média circular. Este valor varia entre 0 e 1, onde 0 representa uma dispersão completa das direções (dados uniformemente distribuídos) e 1 indica que todas as direções são idênticas (dados completamente concentrados em uma única direção). Para o nosso estudo o comprimento do vetor médio foi calculado como 0.775. Esse valor próximo de 1 sugere uma alta concentração das direções dos ventos, indicando que os ventos durante o período analisado vieram predominantemente de direções semelhantes, com pouca dispersão. A variância circular dos dados foi calculada como 0.224 que sugere que as direções do vento são bastante consistentes e estão concentradas em torno da média e o Desvio Padrão dos dados foi calculado como 0.713 e sugere que, embora haja alguma variação nas direções do vento, esta variação não é muito ampla, reforçando a consistência observada nos dados. Essas estatísticas são resumidas na Tabela 5.1.

Tabela 5.1 – Estatísticas Circulares

| **Estatística** | **Valor** |
|---|---|
| Média Circular (graus) | 178.1377 |
| Mediana Circular (graus) | 178 |
| Comprimento do Vetor Médio | 0.7753 |
| Variância Circular | 0.2247 |
| Desvio Padrão Circular | 0.7134 |

Como não é prudente concluir acerca da uniformidade das direções de vento apenas com base nos dados e gráficos visuais, foram empregados testes estatísticos apropriados: o teste de Rayleigh e o teste de Watson. Os resultados desses testes são apresentados na Tabela 5.2, onde são analisados os níveis de concentração direcional e a distribuição das direções de vento.

Tabela 5.2 – Resultados dos Testes de Uniformidade

| **Parâmetros** | **Rayleigh** | **Watson** |
|---|---|---|
| **Estatística** | 0.7753 | 200.8093 |
| **P-valor** | 0 | $< 0.01$ |

O teste de Rayleigh obteve uma estatística de 0.7753, com um P-valor de 0, indicando forte evidência contra a hipótese nula de uniformidade, ou seja, as direções dos ventos não são uniformemente distribuídas. O teste de Watson também avalia a uniformidade, porém é mais sensível a distribuições bimodais. Nossos resultados mostram um teste estatístico de 200.8093, com um P-valor significativamente menor que 0.01, confirmando que as direções dos ventos não seguem uma distribuição uniforme.

### Resultados do Modelo de Regressão Circular para Temperatura Máxima

O modelo de regressão relaciona a **direção do vento** $\theta$ (a variável circular) à **temperatura máxima** $T$ (a variável linear) utilizando uma distribuição von Mises. Obtivemos através do pacote *lm.circular* os seguintes paramêtros:

Tabela 5.3 – Resultados do modelo de regressão circular

| **Parâmetro** | **Valor** |
|---|---|
| Parâmetro de Concentração ($\kappa$) | 2.591724 |
| Erro Padrão de $\kappa$ (se.kappa) | 0.047726 |
| Coeficiente de Regressão ($\beta$) | 0.9194324 |
| Covariância do Coeficiente (cov.coef) | 0.01198991 |
| Erro Padrão do Coeficiente (se.coef) | 0.1094984 |
| Log-Verossimilhança (log.lik) | 3248.62 |
| Valor-t ($t$) | 8.396764 |
| Valor-p ($p$) | 0 |

Fonte: Do Autor

O modelo é definido da seguinte forma:

$$\theta \mid T \sim \mathrm{Von\,Mises}(\mu, \kappa)$$

em que:

$$\mu = \beta \cdot T = 0.9194324 \times T$$
$$\kappa = 2.591724$$

A função de densidade de probabilidade da direção do vento $\theta$ dado a temperatura máxima $T$ é:

$$f(\theta \mid T) = \frac{1}{2\pi I_0(2.591724)} \exp\left(2.591724 \cos\left(\theta - 0.9194324 \times T\right)\right)$$

onde $I_0(2.591724)$ é a função de Bessel modificada de primeira espécie de ordem 0 avaliada em $\kappa = 2.591724$.

**Equação Final do Modelo**

$$\boxed{f(\theta \mid T) = \frac{1}{2\pi I_0(2.591724)} \exp\left(2.591724 \cos\left(\theta - 0.9194324 \times T\right)\right)}$$

Além dos parâmetros do modelo, calculamos métricas de erro para avaliar a precisão das previsões:

- **Erro Médio**: 151.73 graus
- ***RMSE (Root Mean Square Error)***: 154.07 graus

**Interpretação dos Resultados**

A Tabela 5.3 apresenta os resultados do modelo de regressão circular ajustado. A seguir, são detalhadas as interpretações de cada parâmetro estimado:

- **Parâmetro de Concentração ($\kappa$)**: O parâmetro de concentração da distribuição de von Mises é 2.591724. Valores maiores de $\kappa$ indicam que os dados estão mais concentrados em torno da direção média $\mu$.
- **Erro Padrão de $\kappa$ (se.kappa)**: O erro padrão da estimativa de $\kappa$ é 0.047726, indicando a precisão dessa estimativa.

- **Coeficiente de Regressão** ($\beta$): O coeficiente de regressão é 0.9194324, indicando a relação entre a variável explicativa e a variável de resposta circular. Este coeficiente sugere o quanto a variável explicativa influencia a direção da variável de resposta.

- **Covariância do Coeficiente (cov.coef)**: A covariância do coeficiente de regressão é 0.01198991, fornecendo uma medida da variabilidade conjunta dos coeficientes.

- **Erro Padrão do Coeficiente (se.coef)**: O erro padrão do coeficiente de regressão é 0.1094984, indicando a precisão da estimativa do coeficiente.

- **Log-Verossimilhança (log.lik)**: O valor da log-verossimilhança do modelo é 3248.62. Um valor maior indica um melhor ajuste do modelo aos dados.

- **Valor-t** ($t$): O valor-t dos coeficientes da regressão é 8.396764, utilizado para testar a hipótese nula de que o coeficiente é igual a zero. Um valor alto indica que o coeficiente é significativamente diferente de zero.

- **Valor-p** ($p$): O valor-p associado ao valor-t é 0, indicando que o coeficiente de regressão é estatisticamente significativo, geralmente considerado significativo se $p < 0.05$.

Os valores de **Erro Médio** e **RMSE** obtidos são excepcionalmente elevados, indicando que pode não existir uma relação entre a temperatura máxima ($T$) e a direção dos ventos ($\theta$). Um fator que pode estar influenciando é que o modelo circular-linear aplicado assume uma relação linear entre a temperatura máxima e a direção média do vento. Se a verdadeira relação for não linear ou influenciada por outros fatores, o modelo pode não capturar essa complexidade.

### 5.2.6 Resultados do Modelo de Regressão Circular para Velocidade do Vento

O modelo de regressão relaciona a **direção do vento** $\theta$ (uma variável circular) à **velocidade do vento** $V$ (uma variável linear) utilizando uma distribuição von Mises. Obtivemos, através do pacote *lm.circular*, os seguintes parâmetros:

Tabela 5.4 – Resultados do modelo de regressão circular para Velocidade do Vento

| **Parâmetro** | **Valor** |
|---|---|
| Parâmetro de Concentração ($\kappa$) | 2.588748 |
| Erro Padrão de $\kappa$ (se.kappa) | 0.04766412 |
| Coeficiente de Regressão ($\beta$) | 85.09554 |
| Covariância do Coeficiente (cov.coef) | 1221.35 |
| Erro Padrão do Coeficiente (se.coef) | 34.94782 |
| Valor-t ($t$) | 2.434931 |
| Valor-p ($p$) | 0.007447313 |

Fonte: Do Autor

O modelo é definido da seguinte forma:

$$\theta \mid V \sim \text{Von Mises}(\mu, \kappa)$$

em que:

$$\mu = \beta \cdot V = 85.09554 \times V$$
$$\kappa = 2.588748$$

A função de densidade de probabilidade da direção do vento $\theta$ dado a velocidade do vento $V$ é:

$$f(\theta \mid V) = \frac{1}{2\pi I_0(2.588748)} \exp\left(2.588748 \cos\left(\theta - 85.09554 \times V\right)\right)$$

no qual $I_0(2.588748)$ é a função de Bessel modificada de primeira espécie de ordem 0 avaliada em $\kappa = 2.588748$.

**Equação Final do Modelo**

$$\boxed{f(\theta \mid V) = \frac{1}{2\pi I_0(2.588748)} \exp\left(2.588748 \cos\left(\theta - 85.09554 \times V\right)\right)}$$

Além dos parâmetros do modelo, calculamos métricas de erro para avaliar a precisão das previsões:

- **Erro Médio**: 99.66 graus

- ***RMSE (Root Mean Square Error)***: 107.45 graus

**Interpretação dos Resultados**

A Tabela 5.4 apresenta os resultados do modelo de regressão circular ajustado. A seguir, são detalhadas as interpretações de cada parâmetro estimado:

- **Parâmetro de Concentração** ($\kappa$): O parâmetro de concentração da distribuição de von Mises é 2.588748. Valores maiores de $\kappa$ indicam que os dados estão mais concentrados em torno da direção média $\mu$.
- **Erro Padrão de $\kappa$ (se.kappa)**: O erro padrão da estimativa de $\kappa$ é 0.04766412, indicando a precisão dessa estimativa.
- **Coeficiente de Regressão** ($\beta$): O coeficiente de regressão é 85.09554, indicando a relação entre a variável explicativa e a variável de resposta circular. Este coeficiente sugere o quanto a velocidade do vento $V$ influencia a direção da variável de resposta $\theta$.
- **Covariância do Coeficiente (cov.coef)**: A covariância do coeficiente de regressão é 1221.35, fornecendo uma medida da variabilidade conjunta dos coeficientes.
- **Erro Padrão do Coeficiente (se.coef)**: O erro padrão do coeficiente de regressão é 34.94782, indicando a precisão da estimativa do coeficiente.
- **Valor-t** ($t$): O valor-t dos coeficientes da regressão é 2.434931, usado para testar a hipótese nula de que o coeficiente é igual a zero. Um valor acima de 2 geralmente indica que o coeficiente é significativamente diferente de zero.
- **Valor-p** ($p$): O valor-p associado ao valor-t é 0.007447313, indicando que o coeficiente de regressão é estatisticamente significativo, já que $p < 0.05$.

Os valores de **Erro Médio** (99.66 graus) e **RMSE** (107.45 graus) são elevados, embora menores do que os obtidos no modelo anterior que utilizava a temperatura máxima como variável independente. Isso indica que o modelo que considera a velocidade do vento apresenta um desempenho melhor na predição da direção dos ventos, mas ainda enfrenta desafios significativos em termos de precisão.

# 6 CONCLUSÃO

Diante do exposto, os métodos pertencentes à Estatística Circular permitiram a avaliação da direção dos ventos em Campina Grande durante o primeiro semestre de 2022. Este estudo utilizou técnicas específicas para dados circulares, como o cálculo da média circular, mediana circular, variância e desvio padrão circular, considerando a periodicidade natural desses dados, técnicas essas que se diferenciam significativamente dos métodos tradicionais de Estatística.

A análise dos dados revelou uma predominância de ventos oriundos do leste, com uma concentração significativa entre 135° e 175°. A média circular foi determinada em aproximadamente 178.14 graus, indicando que os ventos predominantes foram direcionados levemente ao leste, um padrão influenciado por características geográficas e climáticas da região. Medidas de dispersão como o comprimento do vetor médio (0.775), variância circular (0.224) e desvio padrão circular (0.713) indicaram uma alta concentração das direções dos ventos, com pouca dispersão em torno da média circular.

Para verificar a uniformidade das direções dos ventos, foram aplicados testes estatísticos apropriados. O Teste de Rayleigh e o Teste de Watson foram empregados para avaliar se as direções dos ventos seguem uma distribuição uniforme ao longo do círculo. O Teste de Rayleigh, com uma estatística de 0.7753 e um p-valor de 0, indicou forte evidência contra a hipótese nula de uniformidade, sugerindo que as direções dos ventos são concentradas e não uniformemente distribuídas. O Teste de Watson, com uma estatística de 200.8093 e um p-valor significativamente menor que 0.01, também confirmou que as direções dos ventos não seguem uma distribuição uniforme.

Os modelos de regressão circular aplicados à temperatura máxima e à velocidade do vento demonstraram robustez, com coeficientes de regressão significativos e altos valores de concentração, indicando que os dados estavam fortemente agrupados em torno da direção média. No entanto, as métricas de erro associadas a esses modelos revelaram desafios significativos. O modelo que utilizou a temperatura máxima como variável independente apresentou um Erro Médio de 151.73 graus e um *RMSE (Root Mean Square Error)* de 154.07 graus. Embora o modelo que considerou a velocidade do vento tenha mostrado uma melhoria, com um Erro Médio de 99.66 graus e um RMSE de 107.45 graus, ambos os modelos ainda exibem erros consideráveis, indicando que a relação entre as variáveis independentes e a direção dos ventos não está sendo capturada de maneira satisfatória.

Esses altos valores de Erro Médio e RMSE sugerem que existem fatores adicionais ou relações mais complexas que não foram devidamente incorporados nos modelos atuais. Uma possível causa é a suposição de uma relação linear entre as variáveis independentes (temperatura máxima e velocidade do vento) e a direção média do vento. Se a verdadeira relação for não linear ou influenciada por outras variáveis climáticas, os modelos lineares utilizados podem não estar adequados para capturar essa complexidade. Além disso, a presença de *outliers* ou dados anômalos pode estar distorcendo os resultados, aumentando os erros de predição. A dispersão elevada das direções dos ventos, mesmo com um parâmetro de concentração relativamente alto ($\kappa = 2.591724$), também contribui para a dificuldade dos modelos em prever com precisão as direções observadas.

Para abordar esses desafios, recomenda-se uma revisão das funções de cálculo utilizadas, especialmente a função de distância angular, para garantir que os erros estejam sendo calculados corretamente dentro do intervalo [0°, 180°]. Além disso, explorar modelos que permitam relações não lineares ou a inclusão de múltiplas variáveis explicativas pode melhorar a capacidade preditiva dos modelos. A análise e tratamento de *outliers*, bem como o aumento da amostra de dados, são passos cruciais para assegurar a robustez e a representatividade dos modelos. Técnicas de validação cruzada também devem ser empregadas para avaliar a generalização dos modelos e evitar *overfitting*.

A aplicação dessas técnicas revelou-se crucial para a interpretação precisa dos padrões de vento, que podem ter implicações práticas em diversas áreas, como planejamento urbano, agricultura e energia eólica. A compreensão detalhada dos padrões de vento em Campina Grande pode informar políticas públicas e estratégias de gestão ambiental, contribuindo para o desenvolvimento sustentável da região. Embora os modelos atuais apresentem limitações, os resultados fornecem uma base sólida para análises futuras, permitindo comparações com dados de anos subsequentes para identificar tendências sazonais e mudanças significativas ao longo do tempo.

## REFERÊNCIAS

ABRAMOWITZ, Milton; STEGUN, Irene A. **Handbook of Mathematical Functions**. New York, 376.

BORRADAILE, Graham J. **Statistics of earth science data: their distribution in time, space, and orientation**. Springer, 2003.

FISHER, Nicholas I. **Statistical analysis of circular data**. Cambridge University Press, 1995.

FISHER, Nicholas I.; LEE, Alan J. **Regression models for an angular response**. Biometrics, p. 665-677.

FISHER, Nicholas I.; LEWIS, Toby; EMBLETON, Brian JJ. **Statistical analysis of spherical data**. Cambridge University Press.

HILL, G. W. **New approximations to the von Mises distribution**. Biometrika, p. 673-676.

HILL, G. W. **Algorithm 518: Incomplete Bessel Function I0. The Von Mises Distribution**. ACM Transactions on Mathematical Software (TOMS), v. 3, n. 3, p. 279-284.

JAMMALAMADAKA, S. Rao; SENGUPTA, Ambar. **Topics in circular statistics**. World Scientific, 2001.

JEFFREYS, Harold. **Theory of Probability, 2nd ed**. Oxford University Press, Oxford, 302.

KENT, John T. **Limiting Behaviour of the von Mises-Fisher distribution**. Proc. Cambridge Phil. Soc. 84, 531-536.

MARDIA, Kanti V. **Statistics of directional data**. Academic Press, 1972.

MARDIA, Kanti V.; JUPP, Peter E. **Directional statistics**. Wiley, 2000.

MARDIA, K. V. **Distribution theory for the von Mises-Fisher distribution and its application**. Springer Netherlands. p. 113-130.

PEWSEY, Arthur; NEUHÄUSER, Markus; RUXTON, Graeme D. **Circular statistics in R**. OUP Oxford, 80.

STEPHENS, M. A. **Random Walk on a Circle**. Biometrika 50, 385-390.

STOLL, Clifford, citado em: Keeler, Mark R. **Nothing to Hide: Privacy in the 21st Century**. iUniverse, 2006.

UPTON, G. J. G. **New approximations to the distribution of certain angular statistics**. Biometrika, v. 61, n.2, p. 369-373.

VON MISES, Richard. **Die Wahrscheinlichkeitsrechnung und ihre Anwendungen in der Statistik und in der Theorie der Fehler**.

VON MISES, Richard. **Über die "Ganzzahligkeit"der Atomgewichte und verwandte Fragen**. Physikalische Zeitschrift 19, 1918, 490.

# APÊNDICE A – Código no R

```r
setwd("E:/Faculdade/TCC")

require(circular)
require(cplots)
library(plotrix)

dados <- read.csv2("Dados Metereologicos jan 2022 a jun 2022.csv")

# Contando valores ausentes em cada coluna
na_counts <- sapply(dados, function(x) sum(is.na(x)))

# Exibindo o resultado
print(na_counts)


circ_data <- dados$Dir..Vento..m.s.
circ_data <- circular(circ_data, type = "angles", units = "degrees")

# Média Circular
media_circular <- mean.circular(circ_data)
print(paste("Média Circular:", media_circular))

# Mediana Circular
mediana_circular <- median.circular(circ_data)
print(paste("Mediana Circular (graus):", mediana_circular))

# Comprimento do Vetor Médio
vetor_medio <- rho.circular(circ_data)
print(paste("Comprimento do Vetor Médio:", vetor_medio))

# Variância e Desvio Padrão
variancia_circular <- var.circular(circ_data)
desvio_circular <- sd.circular(circ_data)
print(paste("Variância Circular:", variancia_circular))
print(paste("Desvio Padrão Circular:", desvio_circular))

```

```
# Calcular a densidade circular
densidade <- density.circular(circ_data, bw = 75)

par(mar = c(1,0,1,0)) # Set margin to have better layout
plot(circ_data, cex = 1.5, bin = 720, stack = TRUE, sep = 0.011,
    shrink = 1.3,
xlim = c(-4, 1), ylim = c(-0.1, 0.1))
lines(densidade, col = "red", lwd = 2)
rose.diag(circ_data, bins = 10, col = "skyblue", border = "black",
ticks = TRUE, add = TRUE, tcl.text = -100, show.text = FALSE)

# Teste de Rayleigh
rtest <- rayleigh.test(circ_data)
print(rtest)
print(paste("Estatística Rayleigh:", rtest$statistic))
print(paste("Valor-p Rayleigh:", rtest$p.value))

# Teste de Watson
wtest <- watson.test(circ_data)
print(wtest)
print(paste("Estatística Watson:", wtest$statistic))

# Regressão Circular-Linear
temp_max <- dados$Temp..Max...C.
vel_wind <- dados$Vel..Vento..m.s.

model <- lm.circular(type = "c-l", y = circ_data, x = temp_max, init =
     1, verbose = TRUE)
summary(model)

model2 <- lm.circular(type = "c-l", y = circ_data, x = vel_wind, init
    = 1, verbose = TRUE)
summary(model2)

# Novos dados de Temperatura Máxima (C)
new_temp_max <- c(22.6, 22.2, 22, 21.3, 21.1, 20.8, 20.7, 20.8, 20.7,
    20.8,
22.1, 22.4, 23, 23.4, 23.5, 25.3, 26.2, 26.9, 27.5, 26.5,
25.6, 24.8, 23.9, 22.9)
```

```r

# Novos dados de Velocidade do Vento (m/s) - utilizados na previsão baseada em velocidade
new_vel_wind <- c(1.3, 1.6, 1.3, 0.7, 1.2, 0.9, 0.1, 0.2, 0.5, 0.1,
0.9, 1.4, 1, 0.7, 0.8, 1.1, 2.5, 0.7, 2.5, 3.3,
2, 3.6, 3.9, 3.1)

# Direções do Vento Observadas para os novos dados
new_wind_dir_observed <- c(174, 176, 175, 176, 177, 169, 177, 175, 175, 174,
177, 181, 173, 177, 180, 181, 169, 180, 358, 203,
181, 185, 195, 171)

# Função para ajustar ângulos ao intervalo [0, 360) graus
adjust_angle <- function(angle) {
  angle %% 360
}

# Função para calcular a direção prevista com base na Temperatura Máxima
predict_wind_direction <- function(T, beta = 0.9194324) {
  mu <- beta * T
  mu_adjusted <- adjust_angle(mu)
  return(mu_adjusted)
}

# Função para calcular a direção prevista com base na Velocidade do Vento
predict_wind_direction2 <- function(V, beta = 85.09554) {
  mu <- beta * V
  mu_adjusted <- adjust_angle(mu)
  return(mu_adjusted)
}

# Calcular as direções previstas para os novos dados de temperatura máxima
mu_pred_model1 <- predict_wind_direction(new_temp_max)

```

```r
# Calcular as direções previstas para os novos dados de velocidade do vento
mu_pred_model2 <- predict_wind_direction2(new_vel_wind)

# Exibir as direções previstas
print("Previsões - Modelo Temperatura Máxima:")
print(mu_pred_model1)

print("Previsões - Modelo Velocidade do Vento:")
print(mu_pred_model2)

# Função para calcular a distância angular
angular_distance <- function(pred, actual) {
  d <- abs(pred - actual) %% 360
  d <- ifelse(d > 180, 360 - d, d)
  return(d)
}

# Calcular os erros angulares
error_model1 <- angular_distance(mu_pred_model1, new_wind_dir_observed)
error_model2 <- angular_distance(mu_pred_model2, new_wind_dir_observed)

# Calcular métricas de erro
mean_error_model1 <- mean(error_model1)
rmse_model1 <- sqrt(mean(error_model1^2))

mean_error_model2 <- mean(error_model2)
rmse_model2 <- sqrt(mean(error_model2^2))


# Exibir as métricas de erro
print(paste("Modelo Temperatura Máxima - Erro Médio:", round(mean_error_model1, 2), "graus"))
print(paste("Modelo Temperatura Máxima - RMSE:", round(rmse_model1, 2), "graus"))

```

```
print(paste("Modelo Velocidade do Vento - Erro Médio:", round(mean_
    error_model2, 2), "graus"))
print(paste("Modelo Velocidade do Vento - RMSE:", round(rmse_model2,
    2), "graus"))
```

Listing A.1 – Estatística Direcional