CONEXÕES
CIÊNCIA E TECNOLOGIA

DADOS CIRCULARES: TRATAMENTO, ANÁLISE E UMA APLICAÇÃO A DADOS REAIS

FERNANDA VITAL DE PAULA[1], THAFNE SIRQUEIRA CARVALHO[2]

[1]Universidade Federal do Norte do Tocantins - UFNT
[2]Universidade Federal de Brasília - UnB
<fernandavital@uft.edu.br>, <thafne.sirqueira@mail.uft.edu.br>
DOI: 10.21439/conexoes.v18i0.3300

Resumo. A Estatística Circular é um ramo importante da Estatística que tem sido necessário em diversos campos científicos. As técnicas de análise de dados na reta mostram-se inadequadas para o tratamento de dados circulares e assim, novas técnicas de medida de tendência central e dispersão foram desenvolvidas, como o ângulo médio, mediana, variância e desvio padrão. Além disso, há gráficos específicos para representação de dados circulares como o histograma linear, histograma angular, diagrama de rosas e gráfico de dispersão e distribuições circulares importantes para o estudo da distribuição de dados circulares como a von Mises e a Cardioide. O intuito deste trabalho é apresentar conceitos da Estatística Circular com o propósito de divulgar e motivar o uso de suas ferramentas, após um levantamento bibliográfico, além de ilustrar a aplicação de algumas dessas ferramentas por meio de um exemplo real onde as direções das cidades de origem dos alunos do curso de Licenciatura em Matemática da Universidade Federal do Norte do Tocantins serão analisadas. No estudo de tais direções, a estimação dos parâmetros dos modelos von Mises e Cardioide serão estimados computacionalmente pelo método da máxima verossimilhança, via o *software RStudio*.

Palavras-chave: estatística circular; dados circulares; von mises; cardioide.

CIRCULAR DATA: TREATMENT, ANALYSIS AND AN APPLICATION TO REAL DATA

Abstract. Circular Statistics is an important branch of Statistics that has been needed in many scientific fields. Line data analysis techniques are inadequate for treating circular data and, therefore, new techniques for measuring central tendency and dispersion have been developed, such as mean angle, median, variance and standard deviation. In addition, there are specific graphs for representing circular data, such as linear histogram, angular histogram, rose diagram and scatter plot, and important circular distributions for studying the distribution of circular data such as von Mises and Cardioid. The purpose of this work is to present concepts of Circular Statistics with the purpose of disseminating and motivating the use of its tools, after a bibliographic survey, in addition to illustrating the application of some of these tools through a real example where the directions of the cities of origin of the students of the Degree in Mathematics at the Federal University of Norte do Tocantins will be analyzed. In the study of such directions, the estimation of the parameters of the von Mises and Cardioide models will be computationally estimated by the maximum likelihood method, via the RStudio software.

Keywords: circular statistics; circular data; von mises; cardioid.

1 INTRODUÇÃO

A Estatística Circular é um campo da Estatística que se dedica ao estudo de dados direcionais em duas dimensões. Nesse sentido, os dados podem ser representados como ângulos ou como pontos na circunferência do círculo unitário, no sentido horário ou anti-horário, e são expressos em graus ou radianos, a partir de alguma origem. É importante destacar que tais ângulos não são

CC BY NC SA

necessariamente resultados de medições. Existem dados que podem ser convertidos em ângulos como os horários de ocorrência de crimes em um determinado local.

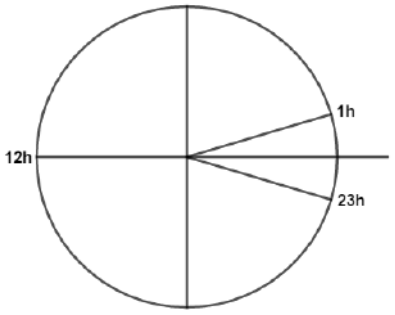
Dados circulares derivam de muitas áreas do conhecimento como Biologia, Geografia e Meteorologia. Alguns exemplos de aplicação são os seguintes:

- Ângulos de ligação das moléculas (Viftrup *et al.*, 2007);
- Direções do vento (Davis; Sampson, 1986);
- Padrões de tempo na incidência de crimes (Brunsdon; Corcoran, 2006);
- Padrões de tempo dos eventos no campo da cibersegurança (Pan; Tomlinson; Koloydenko, 2017);
- Incidência de sarampo ao longo dos anos (Guo; Guo; Li, 2010).

Tratar dados circulares por meio de ferramentas fornecidas pela Estatística usual não é adequado, dadas a sua natureza e periodicidade. Destaca-se que, neste trabalho, o termo Estatística usual será utilizado para se referir às técnicas e aos métodos destinados à análise de variáveis na reta. Um exemplo simples para entender a não adequação das ferramentas estatísticas usuais aos dados circulares pode ser dado pelo seguinte: suponha que exista o interesse em saber o horário em que costumam ocorrer assaltos em um determinado bairro de uma cidade e tem-se uma amostra de tamanho dois com os valores 23h e 01h, ou seja, dois assaltos ocorreram nesses horários. Calculando a média amostral desses dois valores, conforme a Estatística usual, obtém-se 12h. Observando a Figura 1, onde são representados ambos os dados e a média aritmética obtida, observa-se que 12h não é um horário adequado para resumir a amostra.

O exemplo evidencia que ferramentas estatísticas usuais podem não ser suficientes para tratar dados com características periódicas ou direcionais, sendo necessário lançar mão de ferramentas mais apropriadas. Para o caso considerado, a Estatística Circular fornece uma medida denominada direção média circular que tem o mesmo propósito da média aritmética na Estatística usual e é dada por:

Figura 1: Representação dos valores da amostra considerada e da média aritmética usual obtida.



$$\overline{\theta} = \begin{cases} \arctan\left(\frac{y}{x}\right) \text{ quando } x > 0 \text{ e } y \geq 0 \\ \arctan\left(\frac{y}{x}\right) + \pi \text{ quando } x < 0 \\ \arctan\left(\frac{y}{x}\right) + 2\pi \text{ quando } x \geq 0 \text{ e } y < 0 \\ \frac{\pi}{2} \text{ quando } x = 0 \text{ e } y > 0 \\ \text{Indefinido, quando } x = y = 0 \end{cases}, \tag{1}$$

onde

$$x = \frac{\sum_{i=1}^{n} \cos\theta_i}{n} \quad \text{e} \quad y = \frac{\sum_{i=1}^{n} \operatorname{sen}\theta_i}{n}. \tag{2}$$

No exemplo dado, a fim de obter um valor que resuma e represente bem os dados, sob a abordagem da Estatística Circular, o tempo é convertido em ângulos usando a equação

$$\alpha = \frac{360^\circ \cdot h}{24},$$

onde α é a medida angular e h, o horário dado. Nessa conversão, $24h$ equivale a 360° no ciclo trigonométrico e 0° corresponde à $0h$. Para $23h$ e $01h$, tem-se os ângulos 345° e 15°, respectivamente. Assim,

$$x = \frac{\sum_{i=1}^{n} \cos\theta_i}{n} = \frac{\cos 345^\circ + \cos 15^\circ}{2} = 2\cos 15^\circ$$
$$\text{e} \quad y = \frac{\sum_{i=1}^{n} \operatorname{sen}\theta_i}{n} = \frac{\operatorname{sen} 345^\circ + \operatorname{sen} 15^\circ}{2} = 0. \tag{3}$$

Logo,

$$\overline{\theta} = \arctan\left(\frac{y}{x}\right) = \arctan 0 = 0^\circ. \tag{4}$$

Portanto, o horário médio dos assaltos conforme a amostra do exemplo considerado é 0° ou $0h$. Tal valor representa adequadamente a amostra, notando assim que os cálculos da direção média circular e média

arimética se diferenciam e podem fornecer valores distintos.

Neste sentido, este artigo tem o objetivo de apresentar algumas ferramentas fornecidas pela Estatística Circular, por meio de um levantamento bibliográfico. Apesar da grande relevância e possibilidades de aplicações da Estatística Circular, trata-se de um campo carente em referências bibliográficas como manuais, artigos e livros, em português. Tal fato dificulta o conhecimento da potencialidade desse campo pelos pesquisadores brasileiros das áreas mais diversas. Dessa forma, este trabalho vai ao encontro da necessidade de divulgação da Estatística Circular e de motivação ao uso de suas ferramentas. A potencialidade de tais ferramentas é evidenciada pelo tratamento descritivo de dados circulares reais referentes às direções dos municípios de naturalidade dos alunos dos estudantes do Curso de Licenciatura em Matemática da UFNT, campus de Araguaína, e pelo ajuste de modelos probabilísticos aos dados. Para o tratamento dos dados, o pacote *circular* e outras ferramentas do *software RStudio* foram utilizados.

2 FUNDAMENTAÇÃO TEÓRICA

A Estatística Circular se caracteriza pelo tratamento de dados circulares. Tratá-los por meio de ferramentas fornecidas pela Estatística usual não é adequado, conforme evidenciado pelo exemplo apresentado na introdução desse trabalho, tendo em vista a natureza e periodicidade desses dados. Para tratar os dados circulares, "deve-se ter em mente que estes apresentam características próprias, tanto em termos de modelagem quanto em termos de tratamento estatístico descritivo" (Neves, 2015).

A Estatística Circular é um ramo importante da Estatística que tem sido necessário em diversos campos científicos como Biologia, Medicina, Geologia, Meteorologia, entre outros (Paula, 2018). Alguns exemplos são as direções do movimento de um animal após um determinado estímulo, as direções do vento e os horários de chegadas de pacientes em um hospital.

Para a utilização de ferramentas, técnicas e metodologias mais apuradas da Estatística Circular e reconhecimento de suas potencialidades, são necessários alguns conceitos fundamentais que serão apresentados a seguir.

2.1 Conceitos Fundamentais

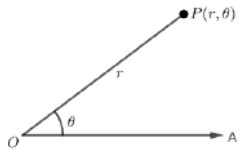
Conforme Leal e Sousa (2003, p. 2), uma observação de uma variável circular pode ser vista como um ponto P sobre o círculo unitário, centrado na origem O do sistema de coordenadas cartesianas e, portanto, tem-se que OP é um vetor unitário em $\mathbb{R}^2$. Essa observação circular pode ainda ser vista como um ângulo θ, formado pelo vetor OP e o semi-eixo Ox.

Ao considerar um ponto P no plano, suas coordenadas podem ser representadas nos sistemas de coordenadas cartesianas ou polares, de modo que as origens de ambos os sistemas são coincidentes. No primeiro, o ponto P é descrito pelo par (x_p, y_p), onde x_p é obtido da projeção de P no eixo x e y_p, da projeção de P no eixo y. Já no sistema polar, P é denotado por (r, θ), onde r é a distância de $(0, 0)$ até P e θ é o ângulo formado no sentido anti-horário entre o semi-eixo positivo Ox e P.

Para converter coordenadas do sistema polar para o cartesiano e vice-versa, utilizam-se os conhecimentos trigonométricos, sobretudo do triângulo retângulo. De acordo com a Figura 2 e considerando que a projeção do ponto P no eixo x determina o ponto X com coordenadas $(x_p, 0)$ e a projeção do ponto P no eixo y determina a ponto Y com coordenadas $(0, y_p)$, obtém-se do triângulo OXP:

$$\operatorname{sen}\theta = \frac{y_p}{r} \Rightarrow y_p = r{\cdot}\operatorname{sen}\theta \quad \text{e} \quad \cos\theta = \frac{x_p}{r} \Rightarrow x_p = r{\cdot}\cos\theta.$$

Figura 2: Representação do ponto P



As equações $x = r \cdot \cos\theta$ e $y = r \cdot \cos\theta$ são válidas para todos os valores de r e θ, pois permitem calcular as coordenadas cartesianas de qualquer ponto quando as coordenadas polares são conhecidas.

Exemplo 2.1 *Encontrar as coordenadas cartesianas do ponto* $\left(4, \frac{\pi}{4}\right)$, *em coordenadas polares.*

Tem-se que $r = 4$ e $\theta = \frac{\pi}{4}$. *Assim,*

$$\begin{aligned} x &= \quad r \cdot \cos\theta = 4 \cdot \cos\left(\frac{\pi}{4}\right) = 4 \cdot \frac{\sqrt{2}}{2} = 2\sqrt{2} \qquad \text{e} \\ y &= \quad r \cdot \sin\theta = 4 \cdot \sin\left(\frac{\pi}{4}\right) = 4 \cdot \frac{\sqrt{2}}{2} = 2\sqrt{2}. \end{aligned}$$

Logo, o ponto, em coordenadas cartesianas, é dado por $\left(2\sqrt{2}, 2\sqrt{2}\right)$.

Agora, para converter coordenadas do sistema cartesiano para o polar, utilizando novamente os conhecimentos trigonométricos e considerando o triângulo retângulo OXP e o teorema de Pitágoras, obtém-se $r = \sqrt{x_p^2 + y_p^2}$ e $\tan\theta = \frac{y_p}{x_p}$.

Como a função tangente é bijetora no intervalo de $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, utiliza-se a função inversa da tangente para determinar o ângulo θ, obtendo $\theta = \arctan \frac{y}{x}$. No entanto, é necessário ter atenção nos sinais de x e y para definir θ.

$$\theta = \begin{cases} \arctan\left(\frac{y}{x}\right) \text{ quando } x > 0 \text{ e } y \geq 0 \\ \arctan\left(\frac{y}{x}\right) + \pi \text{ quando } x < 0 \\ \arctan\left(\frac{y}{x}\right) + 2\pi \text{ quando } x > 0 \text{ e } y < 0 \\ \frac{\pi}{2} \text{ quando } x = 0 \text{ e } y > 0 \\ \text{Indefinido, quando } x = y = 0 \end{cases} . \quad (5)$$

Exemplo 2.2 *Encontrar as coordenadas polares do ponto cartesiano* $(-1, 1)$.

Tem-se que $x = 1$ e $y = -1$. *Assim,*

$$r = \sqrt{x^2 + y^2} = \sqrt{(-1)^2 + 1^2} = \sqrt{2} \text{ e}$$
$$\operatorname{tg}\theta = \frac{y}{x} \Rightarrow \operatorname{tg}\theta = \frac{1}{-1} \Rightarrow \operatorname{tg}\theta = -1.$$

Como $x > 0$ *e* $y < 0$, *então*

$$\theta = \operatorname{arctg}\left(\frac{y}{x}\right) + 2\pi = -\frac{\pi}{4} + 2\pi = \frac{7\pi}{4}.$$

Logo, o ponto em coordenadas polares é $\left(\sqrt{2}, \frac{7\pi}{4}\right)$.

2.2 Estatística Descritiva Circular

Como mostrado anteriormente, é inadequado utilizar medidas da Estatística usual no tratamento de dados circulares. No entanto, descrever variáveis por meio de medidas de posição e dispersão também é essencial no tratamento de dados circulares. Aqui serão apresentadas as medidas descritivas circulares como a direção média, direção mediana, variância e desvio-padrão, relacionadas às ideias de posição, dispersão e concentração dos dados circulares.

2.2.1 Direção média circular

A fórmula para obtenção da direção média circular se distingue da fórmula da média aritmética na Estatística usual, embora as duas tenham a mesma interpretação nos contextos em que são utilizadas. Conforme apresentado na introdução desse artigo, a equação 1 fornece a direção média dos dados circulares (Jammalamadaka; SenGupta, 2001) e a mesma foi exemplificada.

2.2.2 Direção mediana circular

O cálculo da direção mediana circular é similar à mediana na Estatística usual. No contexto circular, a mediana não precisa ser única, podendo assumir qualquer ângulo ϕ para o qual metade dos dados está no arco $[\phi, \phi + \pi)$ e a maioria dos pontos está mais próximo de ϕ do que de $\phi + \pi$. Alguns autores definem a direção mediana circular como a mediana usual (Fisher, 1995).

2.2.3 O Comprimento Médio Resultante

O comprimento médio resultante $\overline{R}$, de modo que $0 \leq \overline{R} \leq 1$, é dado por

$$\overline{R} = \frac{1}{n}\sum_{j=1}^{n} \cos\left(\theta_j - \overline{\theta}\right). \quad (6)$$

O valor de $\overline{R}$ fornece informações sobre o grau de concentração dos ângulos observados. No caso em que todas as observações são iguais, tem-se $\overline{R} = 1$, representando uma concentração máxima dos dados em um único valor e uma variabilidade mínima entre eles. Quando a concentração dos dados é mínima e a variabilidade é máxima, obtém-se $\overline{R} = 0$.

2.2.4 Variância Circular

Outra medida de dispersão circular é a variância circular amostral, definida como

$$V = 1 - \overline{R}.$$

Quanto menor o valor da variância circular, maior a concentração nos dados. Tendo em vista que $0 \leq \overline{R} \leq 1$, então $0 \leq V \leq 1$.

2.2.5 Desvio Padrão Circular

Além da variância, tem-se o desvio padrão circular amostral definido como:

$$\sigma = \{-2log(1 - V)\}^{\frac{1}{2}} = \{-2log\overline{R}\}^{\frac{1}{2}}.$$

Para dados circulares, a definição de desvio-padrão é diferente daquela utilizada para dados lineares, onde o mesmo é dado pela raiz quadrada da variância.

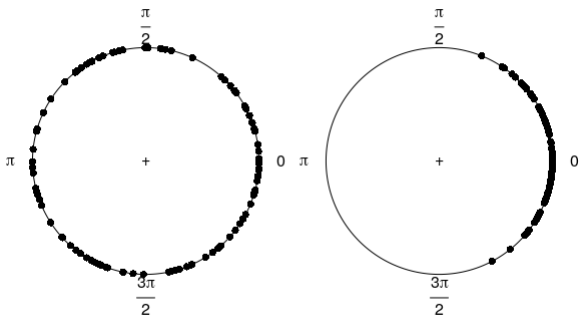
Exemplo 2.3 *A título de ilustração do comportamento dos dados conforme a variância, considere duas amostras de 100 elementos cada, com direção média de* $0°$, *de modo que a primeira possui* $V = 0,988$ *e a outra,* $V = 0,056$. *Plotando os pontos de ambas as amostras no círculo, obtém-se a Figura 3. É possível notar que quando* V *é menor, a concentração dos dados é maior na circunferência à direita.*

2.3 Representação Gráfica

Com relação às representações gráficas de dados circulares, também existem gráficos específicos que podem facilitar a visualização de dados circulares. Neste sentido, pode-se usar os diagramas de rosa, histograma linear, histograma circular e gráficos de dispersão no círculo (Izbicki; Esteves, 2008). Entre eles, descreve-se os principais:

- **Histograma angular**: é obtido envolvendo um histograma linear em torno de um círculo. Cada barra no histograma é centralizada no ponto médio do seu intervalo de agrupamento, com o comprimento da barra proporcional à frequência relativa (ou à quantidade de dados) na classe.

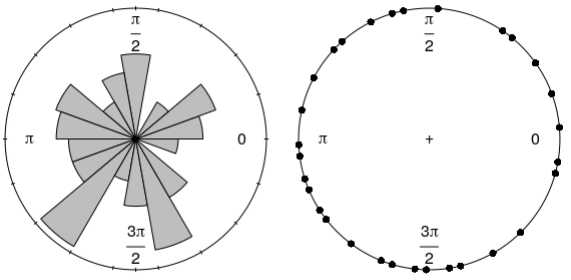
Figura 3: Os dados da circunferência da esquerda apresenta $\mathbf{V = 0,988}$ e da direita, $\mathbf{V = 0,056}$.



- **Diagrama de rosas**: cada grupo é exibido como um setor, de modo que a área do setor é proporcional à frequência do grupo.

Esses gráficos podem ser facilmente construídos através de funções dos pacotes *CircStats* e *Circular* do *RStudio*. Para exemplificação, considera-se as direções em graus tomadas por 100 formigas, presentes em um conjunto de dados denominado *fisherB7* do pacote *Circular*. Por meio de comandos deste pacote, produz-se o diagrama de rosas e o diagrama de dispersão, exibidos na Figura 4.

Figura 4: Diagrama de rosas (à esquerda) e diagrama de dispersão (à direita) dos dados considerados.



2.4 Distribuições Circulares

Uma distribuição circular é uma distribuição de probabilidade na qual a probabilidade está concentrada na circunferência de raio unitário (Neves, 2015). Dada uma variável aleatória circular θ, em radianos, ela pode assumir qualquer valor em intervalo de comprimento 2π, frequentemente sendo $[0, 2\pi)$ ou $[-\pi, \pi)$. No caso contínuo, a função densidade de probabilidade satisfaz às seguintes propriedades:

- $f(\theta) \geq 0$;
- $\int_0^{2\pi} f(\theta)d\theta = 1$;
- $f(\theta) = f(\theta + 2\pi k)$, para qualquer inteiro k.

Existem algumas distribuições circulares mais conhecidas e utilizadas que serão apresentadas neste artigo: Cardioide e von Mises.

2.4.1 Distribuição Cardioide

Considere a variável α, com distribuição Cardioide. Sua função densidade de probabilidade é dada por

$$f(\alpha|\mu,\rho) = \frac{1}{2\pi}\{1 + 2\rho cos(\alpha - \mu)\}, \text{com } 0 \leq \mu < 2\pi \text{ , e } |\rho| < \frac{1}{2}, \quad (7)$$

em que μ é a direção média e ρ é o parâmetro de concentração. Essa distribuição é umimodal e simétrica nas proximidades de μ. Quando $\rho \to 0$, a distribuição converge para a distribuição circular uniforme.

O método de máxima verossimilhança (MMV), criado por Sir Ronald Fisher (1890-1962), também é utilizado para estimação de parâmetros de modelos na Estatística Circular (Pewsey; Neuhäuser; Ruxton, 2013). Lançando mão de seu uso, tem-se

$$L(\mu,\rho|\alpha_1,\ldots\alpha_n) = \prod_{i=1}^{n}\frac{1}{2\pi}\{1 + 2\rho\cos(\alpha_1 - \mu)\} = (2\pi)^{-n}\prod_{i=1}^{n}[1 + 2\rho\cos(\alpha_i - \mu)],$$

resultando na seguinte log-verossimilhança

$$l(\mu,\rho|\alpha_1,\ldots\alpha_n) = \log L(\mu,\rho|\alpha_1,\ldots\alpha_n) = -n\log(2\pi) + \sum_{i=1}^{n}\log\{1 + 2\rho\cos(\alpha_i - \mu)\}.$$

Derivando l em relação a μ e ρ, obtém-se

$$\frac{\partial L}{\partial \mu} = \sum_{i=1}^{n} \frac{1}{1+2\rho\cos(\alpha_i-\mu)} \cdot [1+2\rho\cos(\alpha_i-\mu)]' = \sum_{i=1}^{n} \frac{2\rho\sin(\alpha_i-\mu)}{1+2\rho\cos(\alpha_i-\mu)}.$$

$$\frac{\partial L}{\partial \rho} = \sum_{i=1}^{n} \frac{1}{1+2\rho\cos(\alpha_i-\mu)} \cdot [1+2\rho\cos(\alpha_i-\mu)]' = \sum_{i=1}^{n} \frac{2\cos(\alpha_i-\mu)}{1+2\rho\cos(\alpha_i-\mu)}.$$

Ao igualar ambas as expressões obtidas a 0, para encontrar as estimativas para μ e ρ, observa-se que não existe uma fórmula fechada para as estimativas. Nesse caso, a estimação é realizada computacionalmente, via *RStudio*.

Exemplo 2.4 *Considere um conjunto de dados referentes às direções tomadas por* 76 *tartarugas, presente em um conjunto de dados denominado fisherB3 do pacote Circular. Supondo que o modelo Cardioide seja adequado para ajustar tais direções, estimando os parâmetros ρ e μ computacionalmente no RStudio, obtém-se $\rho = 0,326$ e $\mu = 1,203$.*

2.4.2 Distribuição von Mises

A distribuição von Mises é tão importante na Estatística Circular quanto a distribuição Normal é na Estatística usual. Esse modelo foi proposto pelo físico alemão Richard von Mises, em 1918, ao investigar o desvio de pesos atômicos de valores inteiros (MARDIA; JUPP, 2000).

Seja α uma variável aleatória circular com distribuição von Mises e parâmetros μ e κ, isto é, $\alpha \sim vM(\mu, k)$. Sua função densidade de probabilidade é dada por

$$f(\alpha;\mu,k) = \frac{1}{2\pi I_0(k)} e^{kcos(\alpha-\mu)},\ 0 \leq \alpha < 2\pi,\ 0 \leq \mu < 2\pi \text{ e } k > 0,$$

em que μ é o ângulo médio, κ é o parâmetro de concentração e $I_0(k)$ é a função de Bessel modificada de primeiro tipo e ordem zero, dada pela série

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

quando $k = 0$, a distribuição coincide com a distribuição Uniforme em $[0, 2\pi)$, qualquer que seja o valor de μ.

Aplicando o MMV para estimar os parâmetros da von Mises, tem-se

$$L(\mu, k|\alpha_1, \ldots, \alpha_n) = \prod_{i=1}^{n} \frac{1}{2\pi I_0(k)} e^{kcos(\alpha_i-\mu)} = \frac{1}{[2\pi I_0(k)]^n} e^{\sum k\cos(\alpha_i-\mu)},$$

com log-verossimilhança dada por

$$l(\mu,\kappa|\alpha_1,\ldots,\alpha_n) = \log L(\mu,\kappa|\alpha_1,\ldots,\alpha_n) = -n\ln(2\pi) - nI_0(k) + k\sum\cos(\alpha_i-\mu).$$

Derivando l em relação a μ e κ, obtém-se

$$\frac{\partial L}{\partial \mu} = \kappa \sum \sin(\alpha_i - \mu). \quad (8)$$

$$\frac{\partial L}{\partial \kappa} = -n\frac{1}{I_o(\kappa)} \cdot I_0'(\kappa) + \sum \cos(\alpha_i - \mu). \quad (9)$$

Por meio de manipulações algébricas (Fisher, 1995), igualando as expressões (8) e (9) a zero, obtém-se $\hat{\mu} = \overline{\alpha}$ e $\hat{k} = A^{-1}(\overline{R})$, respectivamente, onde $A(k) = \frac{I_1(k)}{I_0(k)}$ e $I_1(k) = \sum_{r=0}^{\infty}[(r+1)!r!]^{-1}\left(\frac{1}{2}k\right)^{2r+1}$.

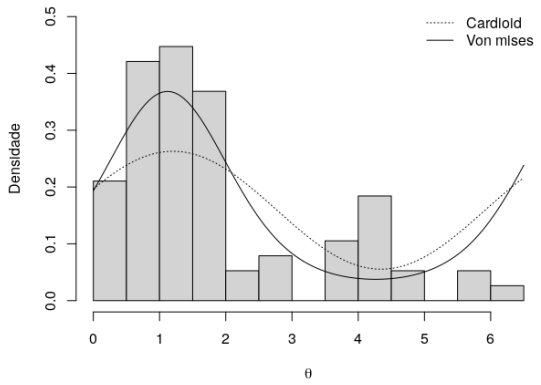
A respeito de $\hat{\mu}$, a distribuição von Mises é a única distribuição circular onde a direção média amostral é o estimador de máxima verossimilhança para a direção média populacional (Bingham; Mardia, 1975).

Exemplo 2.5 *Supondo o conjunto de dados do Exemplo 2.4 com distribuição von Mises, estimando os parâmetros μ e κ no RStudio, obtém-se $\mu = 1,12$ e $\kappa = 1,14$.*

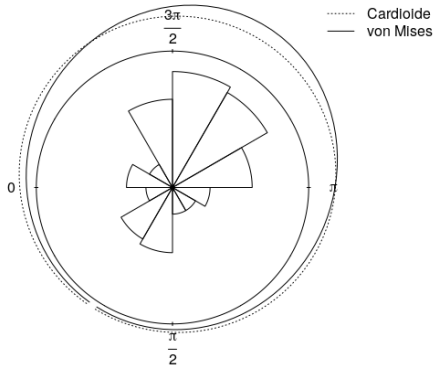
2.4.3 Ajuste das Distribuições Cardioide e Von Mises

O conjunto de dados, utilizados nos Exemplos 2.4 e 2.5, serão representados por meio de histograma usual e de um diagrama de rosas, sendo possível plotar as curvas das distribuições Cardioide e von Mises em ambos os gráficos e visualizar os ajustes, conforme as estimativas encontradas para os parâmetros. A Figura 5 exibe os resultados.

Figura 5: Gráficos dos dados utilizados nos Exemplos 2.4 e 2.5 com os ajustes das distribuições Cardioide e von Mises.



(a) Histograma usual



(b) Diagrama de rosas

3 ESTUDO DIRECIONAL DAS CIDADES DE ORIGEM DOS LICENCIANDOS EM MATEMÁTICA

Nesta seção será feita uma aplicação a dados reais no estudo direcional das cidades de origem dos estudantes do Curso de Licenciatura em Matemática da Universidade Federal do Norte Tocantins (UFNT), campus de Araguaína - TO, matriculados no primeiro semestre de 2020, a fim de ilustrar conceitos e ferramentas da Estatística Circular.

Após a obtenção da naturalidade dos estudantes na Secretaria Acadêmica da UFNT, foram levantadas 63 cidades, como origens dos 256 estudantes do Curso. Inicialmente, para cada uma dessas cidades, foi encontrada sua coordenada geográfica decimal. A Figura 6 exibe um *print* das informações mostradas no *site* ao considerar as cidades de Araguaína e Riachão-TO, a título de exemplificação.

Figura 6: Coordenadas geográficas decimais de Araguaína e Riachão.

Em seguida, por meio das coordenadas decimais obtidas, foram encontradas a direção angular das cidades, considerando Araguaína como referência. Para tal, o endereço <http://pt.thetimenow.com/distance-calculator.php?city1=3465105&city2=3450554> foi utilizado. Considerando as coordenadas das cidades tomadas como exemplo na Figura 6, o *site* exibe as informações dadas na Figura 7.

Figura 7: Direção angular de Riachão em relação à Araguaína.

Como a direção é dada em graus, minutos e segundos pelo site, as direções foram convertidas em graus decimais por meio do aplicativo *A calculadora de graus* cuja interface é exibida pela Figura 8.

Visto que, no período considerado, existiam 256 alunos matriculados no Curso, o conjunto de dados consiste de 256 direções, em graus decimais, referentes às cidades de naturalidade dos alunos. Araguaína foi considerada como cidade de referência, portanto, todas as direções foram encontradas a partir dela. Logo, considerando o círculo, Araguaína está localizada em 0 e as demais cidades, em torno desse valor, considerando o sentido horário.

Figura 8: Conversão da direção para graus decimais.

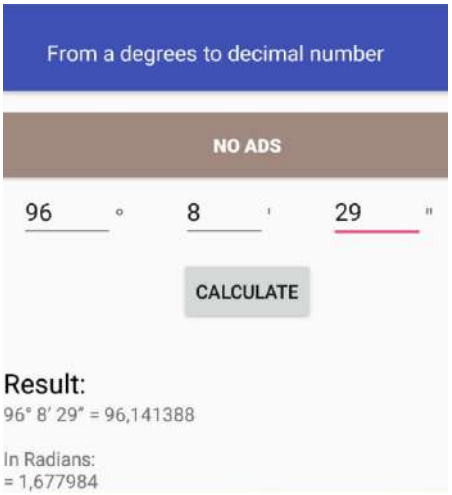
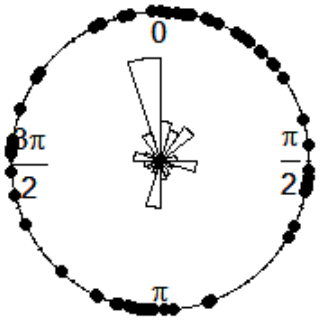


Figura 9: Gráfico de dispersão das direções.



Fonte: próprias autoras.

Após esse procedimento, efetuou-se uma análise descritiva das 256 direções e seguiu-se para o ajuste dos dados a modelos probabilísticos conhecidos, conforme a Estatística Circular, cujos resultados serão apresentados a seguir.

3.1 Análise Descritiva dos Dados

Para o conjunto de direções obtidas, foi utilizado o pacote *circular* do *R-project* para representação dos dados e para uma análise descritiva das direções que permitiu descrever, resumir e compreender os dados por meio de medidas de tendência central e de dispersão.

Vale destacar que, das 256 direções estudadas, 119 correspondem a alunos que nasceram em Araguaína, ou seja, aproximadamente $46,48\%$ dos dados. Além disso, muitas cidades aparecem somente uma vez no conjunto de dados, o que explica o grande número de municípios detectados.

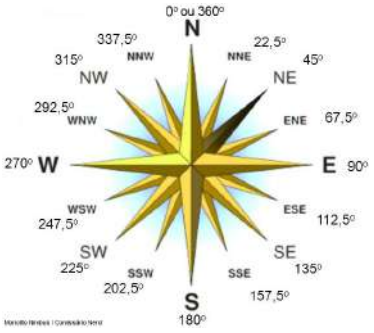
Inicialmente, a Figura 12 exibe o gráfico de dispersão circular dos dados. É importante mencionar que foi adotado o sentido horário para representar os dados. Observa-se que houve uma concentração maior no primeiro quadrante, ou seja, houve uma concentração da origem dos alunos em direções próximas a Araguaína.

Para análise descritiva dos dados, foram encontrados o ângulo médio, a mediana, a moda, o comprimento do vetor médio, a variância circular e o desvio padrão, de modo que os três primeiros correspondem a medidas de posição e os três últimos, a medidas de dispersão. Os valores obtidos são apresentados na Tabela 1, bem como os comandos do pacote *circular* utilizados para obtê-los. Destaca-se que os valores são exibidos em radianos pelo *R-project*.

Pode-se observar que a direção média circular é aproximadamente $0,1177$, que corresponde a $6,7413$. Portanto, a média obtida é um valor que não se distancia de zero, tendo sido influenciada pela maioria dos valores do conjunto de dados ser 0. A média, pode ser localizada na rosa dos ventos, a fim de determinar a direção predominante dos alunos do Curso. Conforme a Figura 10, com rosa dos ventos constituída dos pontos cardeais, colaterais e subcolaterais, a direção média está entre o ponto Norte e Nor-Nordeste.

A média pode ser localizada na Rosa dos ventos, a fim de ser visualizada e melhor compreendida. O valor $6,741307$ denota que a média está entre as direções Norte ($0°$) e Nor-Nordeste ($22,5°$), conforme a Figura 10.

Figura 10: Rosa dos ventos.



Fonte: Monolito, 2021.

De modo prático, como a cidade de referência é Araguaína e $6,7413$ equivale a direção média dos dados, observa-se, na Figura 11, algumas das cidades que estão em torno da média como Araguatins, Augustinópolis, Ananás, Luzinópolis e Açailândia, visto que a direção desses municípios em relação à Araguaína é um valor próximo do ângulo médio.

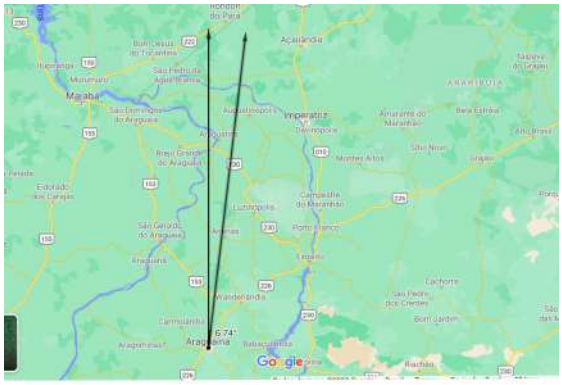
A mediana obtida foi 0, o que significa que 50% dos dados encontra-se em $[0, 180[$ e a moda do conjunto de dados é Araguaína, dado que é a cidade que se repete com maior frequência.

Tabela 1: Medidas de posição e dispersão do conjunto de dados.

Parâmetros	Comandos no pacote circular	Valores
Média circular	mean.circular(x)	0,1177
Mediana circular	median.circular(x)	$0,0000$
Comprimento médio resultante	rho.circular(x)	0,5482
Variância circular	var.circular(x)	0,4518
Desvio padrão circular	sd.circular(x)	1,0964

Fonte: próprias autoras.

Figura 11: Cidades em torno de Araguaína.



Fonte: próprias autoras.

Quanto às medidas de dispersão, a variância resultou em $0,4518$, mais próximo de 0 do que de 1, o que denota que os dados não estão extremamente concentrados em uma certa direção. Concordando com esse resultado, o comprimento médio resultante é $0,5482$. Destaca-se que os dados são mais concentrados quanto mais próximo de zero é o comprimento médio resultante. Visto que esse valor não é tão próximo de 1, existe uma certa dispersão dos dados. Já o desvio-padrão resultou em $1,0964$.

É importante destacar que, embora grande parte dos alunos tenha nascido em Araguaína, isso não significa que eles são de Araguaína. Tendo em vista que a cidade é uma referência no que diz respeito à saúde, muitas pessoas das regiões vizinhas nascem nela, dado que muitas cidades não apresentam estrutura para realização de partos. Assim, considerou-se relevante uma análise descritiva dos dados desconsiderando os dados referentes à Araguaína, ou seja, as direções iguais a 0. Nessa operação, a amostra consistiu-se de 137 alunos que não nasceram em Araguaína. As medidas obtidas, nesse caso, são apresentadas na Tabela 2.

Como direção média dos dados, obteve-se $0,6799$ radianos, o que equivale a $38,95$, aproximadamente. Na rosa dos ventos, da Figura 10, essa média encontra-se entre as direções Nor-nordeste $(22,5)$ e Nordeste (45). Vale ressaltar que, com a retirada de Araguaína, a cidade que mais se repete é Tocantinópolis, sendo a moda do conjunto de dados, tendo como direção $42,53$. Nesse caso, nota-se que a direção média, mais uma vez, é impactada pela moda do conjunto de dados, como no exemplo anterior. Já a mediana circular corresponde a $0,74$.

Tabela 2: Medidas de posição e dispersão do conjunto de dados, sem Araguaína.

Parâmetros	Valores
Média circular	0,6799
Mediana circular	0,7424
Comprimento médio resultante	0,1913
Variância circular	0,8087
Desvio padrão circular	1,8188

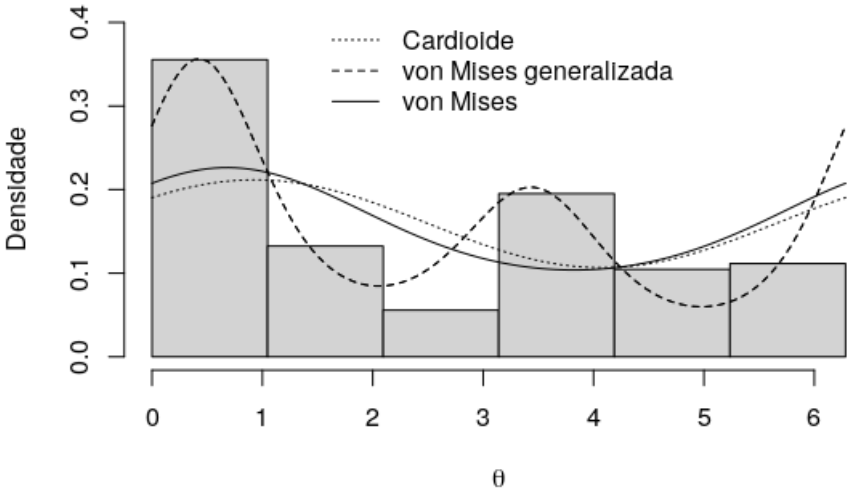
Fonte: próprias autoras.

A Figura 12 exibe a direção média representada no mapa. Observa-se que algumas das cidades que estão em torno da média são Wanderlândia, Darcinópolis, Tocantinópolis, Nazaré e Porto Franco.

Figura 12: Representação da direção média dos dados.

Fonte: próprias autoras.

Figura 13: Histograma dos dados com os ajustes da Cardioide, von Mises e von Mises generalizada.



Fonte: próprias autoras.

No que se refere às medidas de dispersão, o comprimento médio resultante é $0,19$, mais próximo de 0 do que de 1, o que explica a variabilidade dos dados. Como variância obteve-se $0,81$, o que evidencia, também, a pouca concentração dos dados. Já o desvio padrão circular resultou em $1,10$. Comparando a variância com a obtida anteriormente $(0,46)$, observa-se que a dispersão dos dados é ainda maior quando Araguaína é desconsiderada da análise.

3.2 Ajuste de modelos probabilísticos

Nesta seção, modelos circulares probabilísticos serão ajustados ao conjunto de dados considerado, utilizando o MMV. Vale destacar que Araguaína será desconsiderada no ajuste, tendo em vista a justificativa dada na análise descritiva dos dados.

A Figura 13 exibe um histograma dos dados, constituindo-se de 137 direções referentes aos alunos do Curso que não nasceram em Araguaína. É possível observar que trata-se de um conjunto de dados assimétrico e multimodal. Nesse caso, a distribuição von Mises e Cardioide, que foram as distribuições circulares apresentadas neste trabalho, não se adequarão aos dados visto que são distribuições simétricas e unimodais como é possível observar na figura.

Assim, seria mais adequado a esse conjunto de dados um modelo assimétrico e multimodal. A fim de realizar um ajuste mais adequado considerando tais características, uma distribuição assimétrica e multimodal denominada von Mises generalizada foi ajustada aos dados, a fim de exemplificação. A função densidade de probabilidade dessa distribuição é dada por

$$f(\theta|u_1,u_2,k_1,k_2) = \frac{1}{2\pi G_0(\delta,k_1,k_2)} exp\{k_1 cos(\theta-u_1) + k_2 cos2(\theta-u_2)\}, \tag{10}$$

em que $\theta \in [0,2\pi), u_1 \in [0,2\pi), u_2 \in [0,\pi), \delta = (u_1-u_2), k_1 > 0$ e $k_2 > 0$. Além disso,

$$G_0(\delta,k_1,k_2) = \frac{1}{2\pi}\int_0^{2\pi} exp\{k_1 cos\theta + k_2 cos2(\theta-u_2)\}d\theta. \tag{11}$$

Nesse caso, $\theta \sim GvM(u_1, u_2, k_1, k_2)$.

A Figura 13 também exibe a distribuição von Mises generalizada ajustada aos dados. Observa-se que a curva dessa distribuição se adequou melhor ao conjunto de dados que os outros dois modelos. Visto que um estudo mais aprofundado dessa distribuição não se consistiu como objetivo deste trabalho, para um maior conhecimento dessa distribuição, (Gatto; Jammalamadaka, 2007) podem ser consultados.

4 CONSIDERAÇÕES FINAIS

O estudo da Estatística Circular permite fazer a análise descritiva e inferencial de dados com características periódicas ou direcionais, visto que a análise de dados circulares se difere da análise de dados na reta, conforme evidenciado. Neste sentido, no que diz respeito à estatística descritiva, apresentou-se medidas como o ângulo médio, mediana, variância, comprimento médio resultante e desvio padrão circular e representações gráficas como como o diagrama de rosas, dispersão e histograma angular. No que se refere à estatística inferencial, os modelos Cardioide e von Mises foram apresentados bem como as estimativas dos parâmetros de seus modelos por meio do MMV.

Na aplicação apresentada, que consistiu de um estudo direcional das cidades de naturalidade dos estudantes do Curso de Licenciatura em Matemática da UFNT, campus de Araguaína detectou-se que, em média, os estudantes não naturais de Araguaína nasceram em cidades entre as direções nor-nordeste e nordeste, em sua maioria, na cidade de Tocantinópolis. Tendo em vista que os dados analisados são assimétricos e multimodais, os ajustes dos modelos von Mises e Cardioide aos dados não foram adequados. Diante disso, uma distribuição assimétrica e multimodal, denominada von Mises generalizada foi utilizada, a fim de exemplificar que existem outros modelos circulares que podem garantir um bom ajuste aos dados circulares, apesar de não terem sido abordados com maior profundidade neste trabalho.

O campo da Estatística Circular é vasto para pesquisa, principalmente no Brasil, visto que existem poucas referências bibliográficas para pesquisa e estudo em língua portuguesa e promissor, por possuir aplicações em diversas áreas como Biologia, Geografia e Meteorologia. Como exemplo, pode-se citar os ângulos de ligação das moléculas (Viftrup *et al*., 2007), a incidência de sarampo ao longo dos anos (Guo; Guo; Li, 2010) e a direção do vento (Davis; Sampson, 1986). Neste sentido, este trabalho buscou colaborar com esse campo de pesquisa por meio de um aplicação a dados reais que pode auxiliar os interessados a iniciarem seus estudos em Estatística Circular ou a utilizarem suas ferramentas, com o apoio do *software* R-project.

REFERÊNCIAS

ABRAHAO, M. H. M. **A Aventura (Auto) Biográfica:** teoria e empiria. Porto Alegre: EDIPUCRS, 2004.

ABRAMO. **A mulher brasileira nos espaços públicos e privado**. São Paulo: Fundação Perseu Abramo, 2001.

BANCO MUNDIAL. **Relatório sobre o desenvolvimento mundial 1995:** o trabalhador e o processo de integração mundial. Washington: World Bank, 1995.

BINGHAM, M.; MARDIA, K. **Maximum likelihood characterization of the von Mises distribution**. *In*: SPRINGER. A Modern Course on Statistical Distributions in Scientific Work: Volume 3-Characterizations and Applications Proceedings of the NATO Advanced Study Institute held at the University of Calgary, Calgary, Alberta, Canada July 29–August 10, 1974. Canada, 1975. p. 387–398.

BRASIL. **Guia metodológico do sistema de acesso, permanência e êxito**. Brasília: MEC, 2012.

BRASIL. **Guia Metodológico do Sistema de Acesso, Permanência e Êxito.** Brasília: SETEC, 2013. Disponível em: https://x.gd/8Hw9l.

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.

CAVALCANTI, R. M. N. T. **Conceito de cidadania:** sua evolução na educação brasileira a partir da república. Rio de Janeiro: SENAI, 1989. 15 - 52 p.

DAVIS, J. C.; SAMPSON, R. J. **Statistics and data analysis in geology**. New York: Wiley New York, 1986. v. 646.

DINIZ, G.; COELHO, V. **Família e casal: efeitos da contemporaneidade**. *In*: ____. Rio de Janeiro: Ed. PUC-Rio, 2005. cap. A História e as histórias de mulheres sobre o casamento e a família.

FISHER, N. I. **Statistical analysis of circular data**. New York: cambridge university press, 1995.

GATTO, R.; JAMMALAMADAKA, S. R. **The generalized von Mises distribution**. Statistical Methodology, Elsevier, v. 4, n. 3, p. 341–353, 2007.

GUO, H.; GUO, H.; LI, Y. **A study on seasonal distribution of measles in Heping district, Shenyang, from 2001 to 2009**. *In*: IEEE. 2010 3rd International Conference on Biomedical Engineering and Informatics. Yantai, China, 2010. v. 6, p. 2446–2448.

IBGE. **Censo demográfico:** dontagem da população 2010. Rio de Janeiro: IBGE, 2010. Disponível em: www.ibge.gov.br.

IZBICKI, R.; ESTEVES, L. G. **Análise de dados circulares**. IV Simpósio de Iniciação Científica e Pós-Graduação do IME-USP, São Paulo. Atas, p. 69–74, 2008.

JAMMALAMADAKA, S. R.; SENGUPTA, A. **Topics in circular statistics**. Singapore: world scientific, 2001. v. 5.

LACKS, S. **Formação de professores:** a possibilidade da Prática como Articuladora do Conhecimento. Tese (Tese de doutorado) — Faculdade de Educação na Universidade Federal da Bahia, Salvador, 2004.

NEVES, M. R. **Modelos para séries temporais de dados circulares**. Tese (Tese de doutorado) — Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015.

PAN, L.; TOMLINSON, A.; KOLOYDENKO, A. A. **Time pattern analysis of malware by circular statistics**. *In*: IEEE. 2017 ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS). Beijing, China, 2017. p. 119–130.

PAULA, F. V. de. **Extended circular distributions:** mathematical properties, inference and regression model. Tese (Tese de doutorado) — Universidade Federal de Pernambuco, Recife, 2018.

PEWSEY, A.; NEUHÄUSER, M.; RUXTON, G. D. **Circular statistics in R**. United Kington: Oxford University Press, 2013.

SOARES, V. **O Feminismo e o Machismo na percepção das mulheres brasileiras**. *In*: VENTURINI, G.; RECAMAN, M.; OLIVEIRA, S. (Ed.). A Mulher Brasileira nos espaços público e privado. São Paulo: Editora Perseu Abramo, 2004.

VIFTRUP, S. S.; KUMARAPPAN, V.; TRIPPEL, S.; STAPELFELDT, H.; HAMILTON, E.; SEIDEMAN, T. **Holding and spinning molecules in space**. Physical review letters, APS, v. 99, n. 14, p. 143602, 2007.