

Modelagem dos casos de dengue na cidade de Campina Grande-PB usando estatística circular

Divanilda Maia Esteves^{1†}, Débora dos Santos Farias¹

¹Universidade Federal da Paraíba (UFPB), João Pessoa/PB.

Resumo: Este trabalho propõe o uso de estatística circular para modelar o número de casos mensais de dengue na cidade de Campina Grande-PB. A estatística circular é usada para dados que podem ser representados como pontos na circunferência do círculo unitário. O que diferencia esta abordagem da comumente usada é o fato de respeitar a natureza dos dados e fornecer técnicas e gráficos que destacam as características intrínsecas de variáveis direcionais. A principal diferença dessa abordagem é o fato de os dados serem representados por ângulos, o que destaca a natureza periódica dos dados. Os dados utilizados neste trabalho se referem ao número de casos de dengue notificados mensalmente no município de Campina Grande e foram disponibilizados pelo SINAN por meio do TABENET[†]. O período considerado vai de janeiro de 2018 a dezembro de 2021. Os procedimentos estatísticos foram executados no software R. Para as análises, foram usados os pacotes *dplyr* e *Circular*.

Palavras-chave: Estatística Direcional; Dados Circulares; Arbovíroses.

Modeling dengue cases in the city of Campina Grande-PB using circular statistics

Abstract: This work proposes the use of circular statistics to model the number of monthly dengue cases in the city of Campina Grande-PB. Circle statistics is used for data that can be represented as points on the circumference of the unit circle. What sets this approach apart from the commonly used approach is that it respects the nature of the data and provides techniques and graphs that highlight the intrinsic characteristics of directional variables. The main difference of this approach is that the data is represented by angles, which highlights the periodic nature of the data. The data used in this work refer to the number of dengue cases notified monthly in the municipality of Campina Grande and were made available by SINAN through the TABENET[†]. The period considered goes from January 2018 to December 2021. The statistical procedures were performed in software R. For the analyzes the packages *dplyr* and *Circular* were used.

Keywords: Directional Statistics; Circular Data; Arboviruses.

Introdução

Os primeiros registros de uso da estatística circular datam do século IV, onde o Monge Macrobio a utilizou em seus estudos voltados pela astronomia. Em 1858, Florence Nightingale usou um gráfico que foi chamado de *diagrama de rosa* para ressaltar melhorias no saneamento sanitário. No século XIX, Lord Rayleigh (John William Strutt, Terceiro Barão Rayleigh) foi um dos pioneiros no estudo de modelagem de dados circulares, desenvolvendo um teste estatístico formal para uniformidade contra um modelo alternativo de uma única direção preferida (RAYLEIGH, 1880). No entanto, na década de 1950 aconteceram avanços importantes no estudo da estatística circular, como os artigos de Fisher em 1953 (FISHER, 1953) e (WATSON; WILLIAMS, 1956). Nos anos 1960, surgiram outros nomes importantes para o desenvolvimento deste assunto, tais como G. S. Watson, M. A. Stephens e K. W. Mardia. (FISHER, 1995).

[†]Autor correspondente: diana.maia@gmail.com

[†]Tabulador de domínio público desenvolvido pelo DATASUS que permite organizar dados para gerar informações a partir das bases de dados do sistema único de saúde, nele é possível gerar tabelas e produzir gráficos e mapas.

[†]Public domain tabulator developed by DATASUS that allows organizing data to generate information from databases of data from the unified health system, it is possible to generate tables and produce graphs and maps.

Assim, nas últimas décadas a análise de dados circulares vem crescendo como uma importante área da Estatística. Várias novas técnicas surgiram, principalmente para atender às necessidades de muitas áreas do conhecimento, tais como engenharia, medicina, geologia, biologia, geografia, meteorologia, psicologia, oceanografia e matemática, dentre outros.

As direções bidimensionais podem ser representadas como ângulos medidos com respeito a um ponto de partida e um sentido horário ou anti-horário. Como uma direção não tem magnitude, ela pode ser convenientemente representada como pontos na circunferência de um círculo unitário centrado em uma origem ou como vetores unitários conectando a origem a esses pontos. Por motivo desta representação circular, observações em tais direções bidimensionais também são chamados de dados circulares. De forma semelhante os dados direcionais em três dimensões também são referidos como dados esféricos (JAMMALAMADAKA; SENGUPTA, 2001).

A diferença primordial entre o uso da estatística circular e as abordagens clássicas está na representação dos dados em torno de um círculo. Neste caso, cada ponto amostral é associado a um ângulo e isso dá, especialmente na representação gráfica, um destaque maior à natureza periódica da variável considerada. Essa abordagem pode ser usada, entre outras coisas para direção do vento (MAYA-MANZANO et al., 2017), distribuição da ocorrência de eventos ao longo de um dia (BRUNSDON; CORCORAN, 2006) e dados biológicos (LANDLER; RUXTON; MALKEMPER, 2018).

Sendo assim, a presente pesquisa teve como objetivo usar estatística circular para modelar o número de casos de dengue por mês no município de Campina Grande - PB. Para tanto, além de fazer uma análise exploratória dos dados com destaque para a análise gráfica, foi feito um teste para verificar se é plausível a hipótese de que o número de caso se distribui uniformemente ao longo do ano.

Modelagem de Dados Circulares

A Estatística é uma ciência que fornece métodos e princípios para coleta, organização e análise de dados oriundos de estudos ou experimentos, realizados nas mais diversas áreas do conhecimento, com intuito de obtenção de resultados relevantes e assumindo um papel importante no que se refere à tomada de decisão em um acontecimento ou em situações de incertezas (MAGALHÃES; LIMA, 2015). Na Estatística Circular, em particular, cada observação pode ser especificada pelo ângulo formado entre a direção inicial do círculo e o ponto no círculo correspondente à observação, visto que uma direção inicial e uma orientação do círculo sejam determinadas (MARDIA; JUPP, 2000).

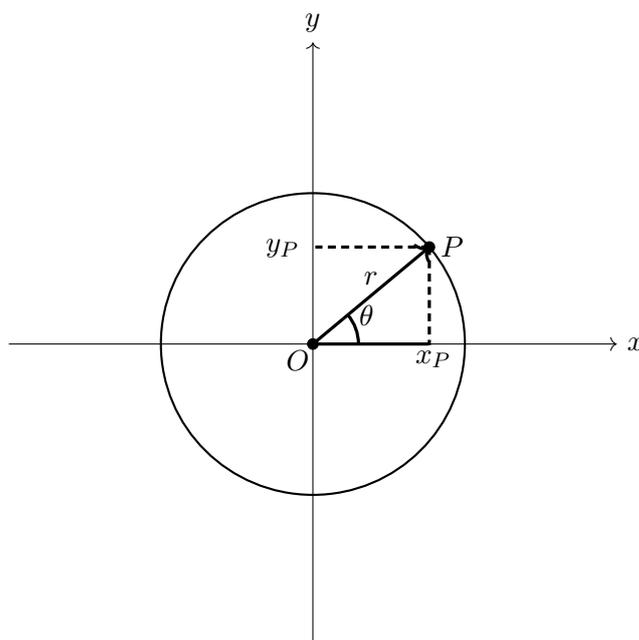
A principal propriedade dos dados circulares é o fato de serem periódicos, ou seja, não há um verdadeiro ponto inicial ou final, mas sim uma escala contínua que retorna em si, ou seja $0 = 2\pi$. Geralmente os dados circulares são medidos em graus. Entretanto, por vezes é mais adequado medi-los em radianos. Lembrando que as medidas angulares são convertidas de graus para radianos multiplicando por $\pi/180$ (MARDIA; JUPP, 2000). Neste caso, essa abordagem dá uma visão diferente da estatística tradicional, porque destaca a essa característica periódica dos dados.

Seja um ponto $P \neq 0$, onde no plano pode ser representado como (x_i, y_i) em termos de coordenadas cartesianas, no qual x_i é a projeção de P no eixo x e y_i , a projeção no eixo y . Pode-se descrever também esse ponto como coordenadas polares (r, θ) , em que r é a distância do ponto até a origem e θ o ângulo tomado no sentido anti-horário, da parte positiva do eixo Ox ao segmento OP , conforme está representado na **Figura 1**.

Essa conversão entre o sistema de coordenadas polares com o sistema de coordenadas cartesianas é feito por meio das funções trigonométricas seno e cosseno, que são dadas por

$$x = r \cos \theta, \quad y = r \sin \theta.$$

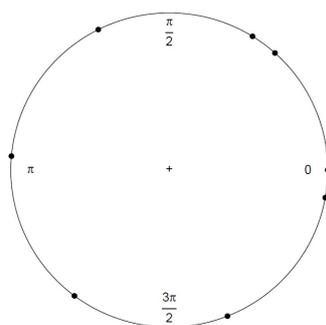
Figura 1: Representação do ponto P em uma unidade circular.



Fonte: Elaborado pelo o autor, 2022.

Um dos pontos fortes do uso da estatística circular é a representação gráfica, que facilita a análise e a interpretação de dados direcionais. Na estatística tradicional, o gráfico de dispersão é feito com os pontos (n, x_n) . No círculo, usam-se os ângulos referentes a cada observação. A **Figura 2** tem uma representação de um gráfico de dispersão circular. O diagrama de rosa é o

Figura 2: Gráfico de dispersão para dados circulares.

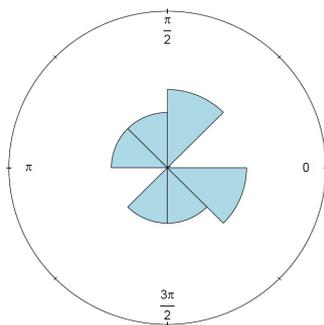


Fonte: Elaborado pelo o autor, 2022.

gráfico mais comumente usado na representação de dados circulares e foi criado por Florence Nightingale (1820-1910). Em sua construção, os raios são proporcionais às frequências de cada classe e, desta forma, há uma dilatação entre a origem e as barreiras dos intervalos (**Figura 3**).

Assim como no caso da representação gráfica, as medidas resumo também tem suas versões para estatística circular. Na média circular não é adequado o uso da média aritmética, porque pode levar a conclusões equivocadas. A média, como também o desvio padrão e outras medidas descritivas sofrem com sua forte dependência na escolha da direção zero e do sentido de rotação. Para ilustrar o cálculo médio do ângulo em dados circulares, suponha duas direções, dados

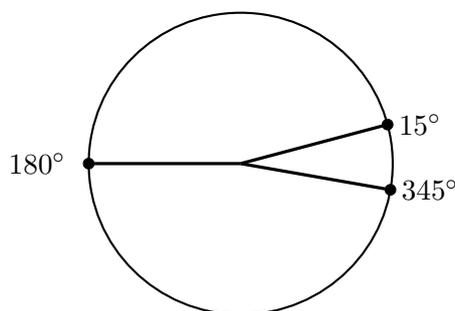
Figura 3: Representação gráfica do diagrama de rosa.



Fonte: Elaborado pelo o autor, 2022.

pelos ângulos 15° e 345° que representam o voo^{iv} dos pássaros e deseja saber a direção média. De acordo a **Figura 4** o valor que corresponde a essa média é 180° , onde aponta para o oeste, enquanto as duas observações apontam essencialmente para o leste. Com esse exemplo fica visível que medidas usadas para variáveis lineares não são adequadas para representar as variáveis circulares. Portanto, é necessário alternativas tanto para medidas de posição como para medidas de dispersão ao lidar com dados circulares (JAMMALAMADAKA; SENGUPTA, 2001).

Figura 4: Representação gráfica da média aritmética em dados circulares.



Fonte: Elaborado pelo o autor, 2022.

Desta forma, suponha n vetores unitários $\vec{OP}_1, \vec{OP}_2, \dots, \vec{OP}_n$, com ângulos correspondentes $\theta_i, i = 1, \dots, n$. Define-se a média circular, $\bar{\theta}$ de $\theta_1, \dots, \theta_n$, correspondente ao vetor resultante da soma $\vec{OP}_1 + \vec{OP}_2 + \dots + \vec{OP}_n$ (MARDIA; JUPP, 2000). Tal vetor é o centro de massa dos pontos $(\cos \theta_1, \sin \theta_1), \dots, (\cos \theta_n, \sin \theta_n)$, e as coordenadas cartesianas do centro de massa são (\bar{C}, \bar{S}) dadas por

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n \cos \theta_i, \quad \bar{S} = \frac{1}{n} \sum_{i=1}^n \sin \theta_i,$$

e a média direcional $\bar{\theta}$ será a solução das equações

$$\bar{C} = \bar{R} \cos \bar{\theta}, \quad \bar{S} = \bar{R} \sin \bar{\theta}.$$

^{iv}As direções dos pássaros: Norte, Sul, Leste e Oeste (são as coordenadas geográficas); leste como direção zero e sentido anti-horário como sentido de rotação positiva.

Considere

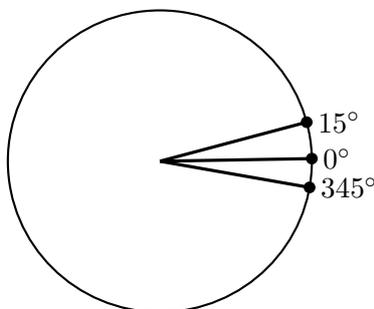
$$\bar{R} = (\bar{C}^2 + \bar{S}^2)^{1/2}$$

como comprimento médio resultante, se $\bar{R} > 0$. Em relação à média direcional, se $\bar{R} = 0$, $\bar{\theta}$ não está definida e se $\bar{R} > 0$, a direção média será

$$\bar{\theta} = \begin{cases} \tan^{-1}(\bar{S}/\bar{C}), & \text{se } \bar{S} \geq 0 \text{ e } \bar{C} > 0, \\ \tan^{-1}(\bar{S}/\bar{C}) + \pi, & \text{se } \bar{C} < 0, \\ \tan^{-1}(\bar{S}/\bar{C}) + 2\pi, & \text{se e } \bar{C} < 0, \\ \text{indefinido}, & \bar{S} = 0, \bar{C} = 0. \end{cases}$$

No exemplo do voo dos pássaros, a direção média circular é o ângulo 0° e isso está representado na **Figura 5**. Percebe-se que o valor está coerente com a interpretação dessa medida de posição.

Figura 5: Representação gráfica da direção média circular em dados circulares.



Fonte: Elaborado pelo o autor, 2022.

A mediana também tem a sua versão adaptada aos dados direcionais (MARDIA; JUPP, 2000).

Além das medidas de posição, a estatística circular usa medidas próprias equivalentes às medidas de dispersão para dados lineares, como comprimento médio do vetor médio resultante e variância circular.

O comprimento médio resultante \bar{R} é utilizado como medida de concentração dos dados e encontra-se no intervalo $(0, 1)$. Seus valores extremos têm algumas propriedades interessantes. Se as direções $\theta_1, \dots, \theta_n$ estão fortemente agrupadas, tem-se $\bar{R} = 1$, representando uma concentração máxima dos dados, e corresponde a uma variabilidade mínima entre eles. Por outro lado, se $\theta_1, \dots, \theta_n$ estejam dispersos no círculo, nesta situação tem-se uma concentração mínima $\bar{R} = 0$ e uma variabilidade máxima. No entanto, $\bar{R} \simeq 0$ não significa que as observações estão dispersas uniformemente ao redor do círculo (MARDIA; JUPP, 2000).

Para propósitos mais descritivos e inferenciais, o comprimento médio resultante \bar{R} é mais importante do que qualquer medida de dispersão, no entanto a sua utilização não é usual. Todavia, para fins de comparação com dados na reta, às vezes é útil considerar medidas de dispersão de dados circulares denotada de variância circular amostral, que pode ser definida por (MARDIA; JUPP, 2000):

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

A definição é análoga à variância de dados lineares, no sentido de quanto menor o valor da variância circular, maior a concentração nos dados. Por construção,

$$0 \leq V \leq 1 \text{ e } 0 \leq \bar{R} \leq 1,$$

sendo que quanto mais próximo de 1, mais dispersos estão os dados e quanto mais próximo de 0, mais concentrados.

O conceito de desvio padrão é diferente do conceito para dados lineares, que equivale à raiz quadrada da variância. O desvio padrão circular amostral é definido por:

$$v = \{-2 \log(1 - V)\}^{1/2} = \{-2 \log \bar{R}\}^{1/2}.$$

Observe que v assume valores em $[0, \infty]$, enquanto V assume valores em $[0, 1]$. Uma boa aproximação para o desvio padrão circular é dada por:

$$v \simeq \begin{cases} (2V)^{1/2}, & \text{para } V \text{ pequeno } (V < 0,18), \\ [2(1 - \bar{R})]^{1/2} & \text{para } \bar{R} \text{ grande } (\bar{R} > 0,82). \end{cases}$$

Uma vez feita a análise exploratória dos dados, o próximo passo seria usar técnicas de inferência para estudar o fenômeno de forma mais global. Uma etapa intermediária consiste em estudar modelos probabilísticos específicos para variáveis com suporte no intervalo $[0, 2\pi]$.

As distribuições circulares são essencialmente de dois tipos: podem ser discretas ou podem ser absolutamente contínuas. No caso contínuo, define-se a função de distribuição como

$$F(x) = Pr(0 < \theta \leq x), \quad 0 \leq x \leq 2\pi$$

$$F(x + 2\pi) - F(x) = 1, \quad -\infty < x < \infty.$$

A partir daí, define-se também a função de densidade de probabilidade $f(\theta)$, a qual possui as seguintes propriedades (FISHER, 1995):

- (i) $f(\theta) \geq 0$, $-\infty < \theta < \infty$;
- (ii) $\int_0^{2\pi} f(\theta) d\theta = 1$;
- (iii) $f(\theta) = f(\theta + 2\pi)$ (f é periódica com período 2π).

A distribuição Uniforme é o modelo mais básico sobre o círculo. Todas as direções são igualmente prováveis. Observe que esta distribuição apresenta como características fundamentais não possuir uma média circular bem definida e o comprimento do vetor resultante ser igual a zero. Assim, a probabilidade total está espalhada uniformemente sobre a circunferência. A função densidade de probabilidade é dada por (JAMMALAMADAKA; SENGUPTA, 2001):

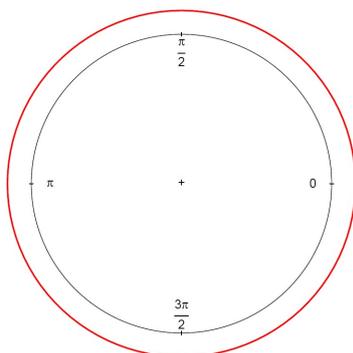
$$f(\theta) = \frac{1}{2\pi}, \quad 0 \leq \theta \leq 2\pi$$

e sua representação gráfica pode ser vista na **Figura 6**.

Quando uma distribuição não é Uniforme, haverá concentração em uma ou mais direções preferenciais. Mais comumente, se for apenas em uma direção a distribuição é chamada de unimodal.

Do ponto de vista da inferência estatística, a distribuição Von Misses é uma das mais importantes para dados circulares. É uma distribuição unimodal simétrica e é o modelo mais comum para amostras unimodais de dados circulares. Ela é tão importante para dados circulares quanto o modelo Gaussiano o é para dados lineares em geral. Esse modelo foi proposto pelo físico alemão R. Von Mises, em 1918, ao investigar desvio de pesos atômicos de valores inteiros (MARDIA; JUPP, 2000). Além da distribuição Von Misses, outros modelos de probabilidade são definidos e podem ser vistos em livros clássicos como (FISHER, 1995) e (MARDIA; JUPP, 2000).

Figura 6: Gráfico da Função de Densidade de Distribuição Uniforme.



Fonte: Elaborado pelo o autor, 2022.

De acordo com (GÓIS, 2002), pode-se verificar que a Estatística Circular dispõe de uma série de testes que avaliam a qualidade do ajustamento das distribuições teóricas referentes aos dados resultantes das observações. A princípio deve-se supor a distribuição principal, no entanto, a busca por métodos mais robustos para uma grande classe de modelos possíveis levam naturalmente para os testes não-paramétricos (PEWSEY, 2013).

O teste de Rayleigh é usado para verificar a distribuição das observações da amostra estudada por toda circunferência. Este método baseia-se no fato do comprimento do vetor médio \bar{R} apresentar um grande valor, significa uma grande concentração dos dados, assim existirá uma menor probabilidade dos dados serem distribuídos de maneira uniforme em torno da média, ou seja, se p-valor for inferior ao nível de significância escolhido (normalmente 0,05) decide-se que os dados não parecem apresentar distribuição Uniforme, indicando uma direção preferencial. A estatística de teste é dada por (MARDIA; JUPP, 2000):

$$Z = 2n\bar{R}^2$$

sendo que n é número de observações e \bar{R} é o comprimento do vetor médio. Sob H_0 , esta estatística de teste tem distribuição χ_2^2 e assim os valores da estatística de teste podem ser usados para encontrar a função do p-valor.

Quando não rejeita-se a hipótese nula, compreende-se que a distribuição circular é uniforme e quando a hipótese alternativa não é descartada, compreende-se que a distribuição circular não é uniformemente distribuída por toda circunferência.

Material e Métodos

Para o desenvolvimento deste trabalho foram utilizados dados reais provenientes do Sistema de Informações de Agravos de Notificação (SINAN) para obtenção de informações sobre os casos de dengue no município de Campina Grande. Os dados analisados compreendem números mensais no período de janeiro de 2018 a dezembro de 2021.

A cidade de Campina Grande está localizada na Mesorregião do Agreste Paraibano, possuindo uma área territorial de 591,658 km^2 , e com população estimada em 413.830 habitantes^v. O clima é caracterizado como tropical semiúmido, com chuvas geralmente de fevereiro a setembro, sendo que o período mais chuvoso ocorre entre maio e julho.

Segundo o Ministério da Saúde, o SINAN foi implantado, a partir de 1993 e a utilização dos dados epidemiológicos permitem que os profissionais de saúde consigam identificar a reali-

^v<https://www.ibge.gov.br/cidades-e-estados/pb/campina-grande.html>

dade epidemiológica de determinadas áreas geográficas, e baseado nisso auxiliar no planejamento da saúde e definir prioridades de intervenção. O sistema é alimentado pela notificação e investigação de casos de doenças e agravos que compõem a lista nacional de doenças de notificação compulsória.

Antes de analisar os resultados foi preciso realizar o tratamento das bases disponibilizada do SINAN por meio do TABENET^{vi}. Sendo assim, o primeiro passo do tratamento da base retirada do SINAN foi efetuar a filtragem referente ao município de notificação, em seguida ao mês de notificação e períodos disponíveis, onde os períodos para análise são relacionados aos anos de 2018 a 2021. A partir dessas manipulações foram criados novos bancos de dados em um editor de planilhas com as seguintes variáveis:

- **mes:** janeiro a dezembro;
- **casosCG:** casos prováveis segundo município de Campina Grande - PB referentes a cada mês;
- **ângulo:** ângulo referente a cada mês.

Como a Estatística Circular se encarrega de estudar o direcionamento dos dados, os valores dos casos prováveis de dengue e os ângulos são números inteiros. Assim, os dados foram associados a um ângulo. Considerando o mês de janeiro como referência, foram utilizados 12 intervalos de 30°. Desta forma, cada mês correspondeu a um valor distribuído ao longo da circunferência.

Os procedimentos estatísticos realizados neste trabalho foram executados no software R (R CORE TEAM, 2021). Para as análises foram utilizados os pacotes `dplyr`, que realiza manipulações de dados, e `Circular` para as análises de dados circulares.

Resultados e discussão

Inicialmente foi feita uma análise exploratória dos dados através de gráficos e medidas resumo para os dados de cada ano.

Na **Figura 7** está representado o diagrama de rosa e o gráfico de dispersão circular para os casos de dengue no ano de 2018. Observa-se que os dados estão mais concentrados nos meses de maio a julho.

Observando o diagrama circular e o gráfico de dispersão circular para os casos de dengue no ano de 2019, que está representado na **Figura 8**, percebe-se uma maior concentração dos dados entre os meses de abril e setembro.

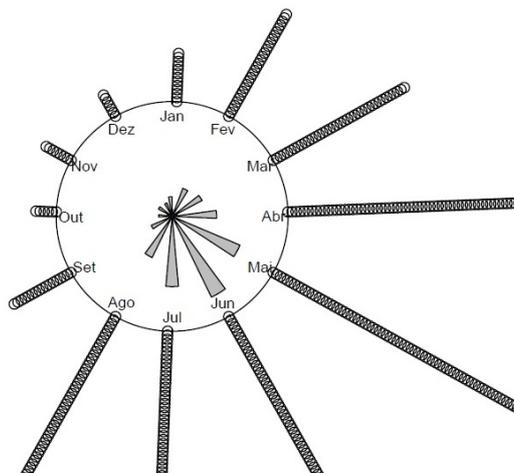
Em relação ao ano de 2020, conforme pode ser visto na **Figura 9**, o diagrama circular e o gráfico de dispersão circular para os casos de dengue no ano de 2020 indicam que os dados se concentram nos meses de maio a setembro.

Para os casos de dengue no ano de 2021, pode-se notar que o diagrama circular e o gráfico de dispersão circular representados na **Figura 10** indicam uma maior concentração dos dados nos meses de maio a novembro.

Para os anos considerados, basendo-se na análise gráfica, há indícios de que os dados sigam uma distribuição unimodal e também de que o número de casos mensais não está uniformemente distribuídos ao longo do ano. Para caracterizar os casos prováveis de dengue na cidade de Campina Grande, usou-se as medidas de posição e as medidas de dispersão que estão apresentadas na **Tabela 1**.

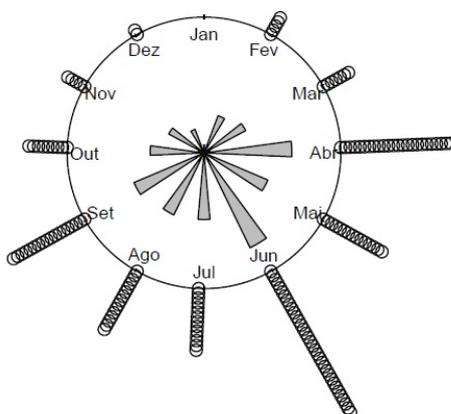
^{vi}http://tabnet.datasus.gov.br/Tutorial/Tutorial_tabNet_FINAL.pdf

Figura 7: Representação do gráfico de dispersão circular e do diagrama de rosa dos casos prováveis de dengue no município de Campina Grande - PB no ano 2018.



Fonte: Elaborado pelo o autor, 2022.

Figura 8: Representação do gráfico de dispersão circular e do diagrama de rosa dos casos prováveis de dengue no município de Campina Grande - PB no ano 2019.



Fonte: Elaborado pelo o autor, 2022.

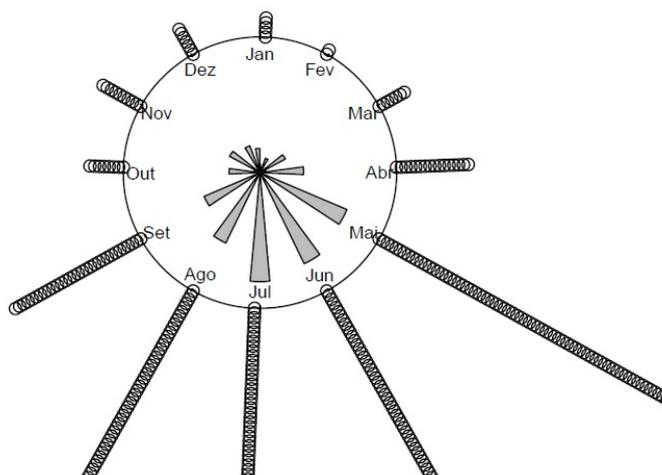
Tabela 1: Dados circulares de casos prováveis de dengue na cidade de Campina Grande - PB.

Ano	n	$\bar{\mu}(^\circ)$	$\bar{\mu}(rad)$	$\tilde{\theta}(^\circ)$	$\tilde{\theta}(rad)$	\bar{R}	V	v
2018	867	145,4592°	2,5387	150°	2,6180	0,6841	0,3159	0,8714
2019	149	157,4519°	2,7481	150°	2,6180	0,4487	0,5513	1,2661
2020	401	166,9832°	2,9144	180°	3,1420	0,6478	0,3522	0,9319
2021	118	185,4058°	-3,0472	180°	3,1420	0,3807	0,6193	1,3898

Fonte: Elaborado pelo o autor, 2022.

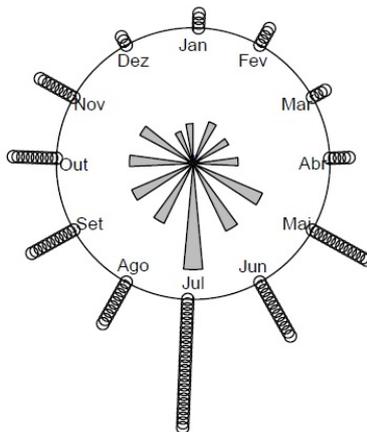
Ao analisar os casos prováveis de dengue na cidade de Campina Grande para o ano de 2018, a direção média foi igual a 2,5387 radianos que corresponde a 145,4592°, para a direção mediana dos dados, o valor foi igual a 2,6180 radianos, correspondendo a direção ao mês de

Figura 9: Representação do gráfico de dispersão circular e do diagrama de rosa dos casos prováveis de dengue no município de Campina Grande - PB no ano 2020.



Fonte: Elaborado pelo o autor, 2022.

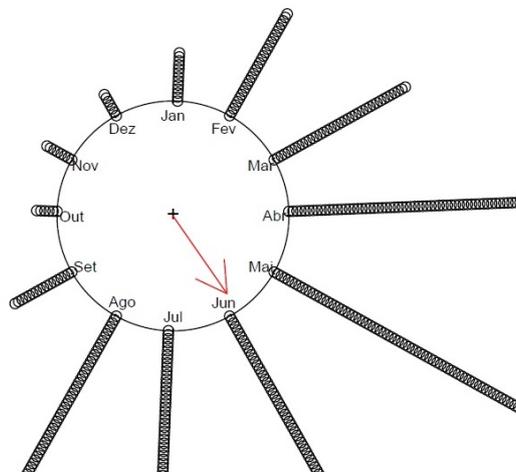
Figura 10: Representação do gráfico de dispersão circular e do diagrama de rosa dos casos prováveis de dengue no município de Campina Grande - PB no ano 2021.



Fonte: Elaborado pelo o autor, 2022.

junho. Para esses dados então, há pouca diferença entre as direções média e mediana. Essa semelhança entre seus valores é consistente com a distribuição dos dados, sendo próxima da simetria. A variabilidade em torno da direção média foi igual a 0,3159, sendo o ano com a menor variabilidade, nota-se que seu valor não foi próximo de 1, o que denota que os dados não estão dispersos na circunferência, também pode ser evidenciado concentração dos dados ao longo da direção média, de acordo os dados da [Tabela 1](#). Desta forma, pode-se constatar ao longo dos 12 meses do ano a direção média dos casos prováveis de dengue foi aproximadamente no mês de junho, conforme está representado na [Figura 11](#).

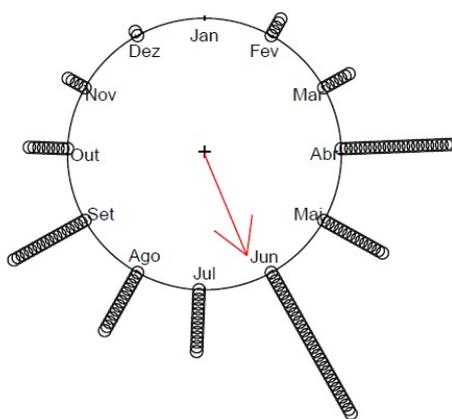
Figura 11: Representação do gráfico de dispersão circular dos casos prováveis de dengue no município de Campina Grande - PB no ano 2018.



Fonte: Elaborado pelo o autor, 2022.

Em relação aos casos prováveis de dengue na cidade de Campina Grande para o ano de 2019, tem como direção média 2,7481 radianos que corresponde a $157,4519^\circ$, a direção mediana foi igual a 2,6180 radianos correspondendo a direção dos dados ao mês de junho, semelhante ao ano anterior. A variabilidade em torno da direção média foi igual 0,5513, observa-se que seu valor não foi próximo de 1, o que evidencia que os dados não estão dispersos na circunferência, também pode ser demonstrado que há uma razoável concentração dos dados ao longo da direção média, em conformidade com os dados da [Tabela 1](#). Assim, pode-se notar ao longo dos 12 meses do ano a direção média dos casos prováveis de dengue em torno do mês de junho, conforme está representado na [Figura 12](#).

Figura 12: Representação do gráfico de dispersão circular dos casos prováveis de dengue no município de Campina Grande - PB no ano 2019.

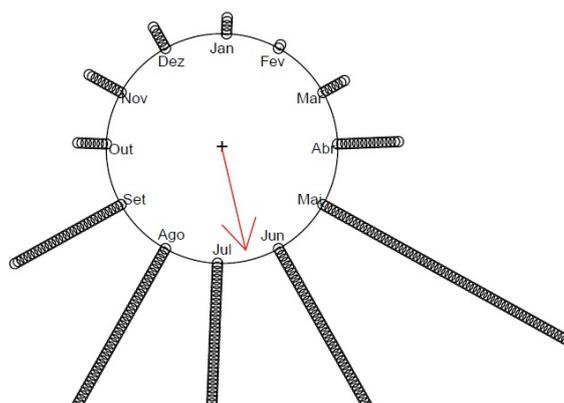


fonte: Elaborado pelo o autor, 2022.

Para o ano de 2020, os casos prováveis de dengue na cidade de Campina Grande teve como direção média de 2,9144 radianos que condiz a $166,9832^\circ$, para direção mediana um valor de 3,1420 radianos, representando uma direção ao mês de julho, onde percebe-se pouca diferença

para esses dados entre a média e a mediana. A variabilidade em torno da direção média igual 0,3522, pode-se observar que seu valor não foi próximo de 1, o que demonstra que os dados não estão dispersos na circunferência, nota-se que houve aglomerações em torno da média, de acordo com os dados da [Tabela 1](#). Sendo assim, constata-se que ao longo dos 12 meses do ano a direção média dos casos prováveis de dengue foi próximo ao mês de julho conforme representado na [Figura 13](#).

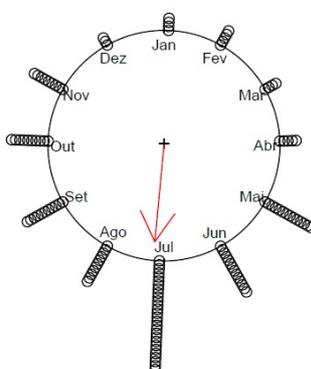
Figura 13: Representação do gráfico de dispersão circular dos casos prováveis de dengue no município de Campina Grande - PB no ano 2020.



fonte: Elaborado pelo o autor, 2022.

A direção média dos casos prováveis de dengue na cidade de Campina Grande no ano de 2021 foi igual a $-3,0472$ radianos que corresponde a $185,4058^\circ$, a direção mediana igual a $3,1420$ radianos, o que corresponde uma direção ao mês de julho, e a variabilidade em torno da direção média igual 0,6193, onde representa uma maior variabilidade dos dados, observa-se que seu valor não foi próximo de 1, o que confirma que os dados não estão dispersos na circunferência, nota-se que houve uma baixa concentração dos dados em torno da média, conforme os dados da [Tabela 1](#). Em vista disso, percebe-se que ao longo dos 12 meses do ano a direção média dos casos prováveis de dengue foi próximo ao mês de julho de acordo a [Figura 14](#).

Figura 14: Representação do gráfico de dispersão circular dos casos prováveis de dengue no município de Campina Grande - PB no ano 2021.



fonte: Elaborado pelo o autor, 2022.

Para realização dos testes de Uniformidade aplicou-se o teste Rayleigh obtendo, em todos os casos, p-valor menor do que 0,001. Isso reitera a hipótese de não uniformidade observada através da análise gráfica.

Conclusão

Diante do exposto, os métodos pertencentes a Estatística Circular permitem a avaliação da concentração dos dados em uma circunferência, como também proporciona o cálculo da média circular, mediana circular, da variância e do desvio padrão circular, levando em consideração a periodicidade natural desse tipo de dados, o que é bem diferente aos dados estudados pela Estatística tradicional.

Também verificou-se tanto por meio do Teste de Rayleigh, de Watson e de espaçamento de Rao a existência de casos prováveis de dengue na cidade de Campina Grande em direções não aleatórias, garantido uma direção específica (sazonalidade dos dados) dos casos de dengue para o período de 4 anos.

Em vista disto, a proposta deste trabalho foi apresentar e desenvolver a metodologia para medidas circulares e avaliar os casos de dengue no município de Campina Grande. Além do mais, os resultados demonstraram que os casos de dengue aumentam no período chuvoso (maio a julho) e que isso deve-se a ocorrência de fatores climáticos, acúmulo de água, falta de apoio de gestão, que contribuem para formação de ambientes favoráveis para o desenvolvimento e a proliferação do vetor.

Sendo assim, vale salientar que a disseminação dessas técnicas é de suma importância para a obtenção adequada dos casos que envolvem as medidas circulares, proporcionando interpretações coerentes dos resultados, quer sejam para representação gráfica, análises descritivas ou mesmo para ajustar um determinado modelo.

Referências

- BRUNSDON, C.; CORCORAN, J. Using circular statistics to analyse time patterns in crime incidence. *Computers, Environment and Urban Systems*, v. 30, p. 300–319, 2006.
- FISHER. Dispersion on a sphere, proc. r. In: *Soc. London*. [S.l.: s.n.], 1953. v. 17.
- FISHER, N. I. *Statistical analysis of circular data*. [S.l.]: cambridge university press, 1995.
- GÓIS, J. E. S. Contribuição dos modelos estocásticos para o estudo da climatologia urbana. 2002.
- JAMMALAMADAKA, S. R.; SENGUPTA, A. *Topics in circular statistics*. [S.l.]: world scientific, 2001. v. 5.
- LANDLER, L.; RUXTON, G. D.; MALKEMPER, E. P. Circular data in biology: advice for effectively implementing statistical procedures. *Behavioral Ecology and Sociobiology*, v. 72, n. 128, 2018.
- MAGALHÃES, M. N.; LIMA, A. C. P. D. *Noções de probabilidade e estatística*. [S.l.]: Editora da Universidade de São Paulo, 2015. v. 7.
- MARDIA, K. V.; JUPP. *Directional statistics*. [S.l.]: Wiley Online Library, 2000. v. 2.
- MAYA-MANZANO, J.; SADYŚ, M.; TORMO-MOLINA, R.; FERNÁNDEZ-RODRÍGUEZ, S.; OTEROS, J.; SILVA-PALACIOS, I.; GONZALO-GARIJO, A. Relationships between airborne pollen grains, wind direction and land cover using gis and circular statistics. *Science of the Total Environment*, v. 584–585, p. 603–613, 2017.

PEWSEY. *Circular statistics in R*. [S.l.]: Oxford University Press, 2013.

R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2021. Disponível em: <http://www.R-project.org/>.

RAYLEIGH, L. On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Taylor & Francis, v. 10, n. 60, p. 73–78, 1880.

WATSON, G. S.; WILLIAMS, E. J. On the construction of significance tests on the circle and the sphere. *Biometrika*, JSTOR, v. 43, n. 3/4, p. 344–352, 1956.