



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA
CURSO DE GRADUAÇÃO EM ESTATÍSTICA

ANNY SUELLEN GOMES DA SILVA

MODELAGEM ESTATÍSTICA PARA DADOS CIRCULARES

FORTALEZA

2019

ANNY SUELLEN GOMES DA SILVA

MODELAGEM ESTATÍSTICA PARA DADOS CIRCULARES

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Estatística do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Estatística.

Orientador: Prof. Dr. Gualberto Segundo Agamez Montalvo

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

D11m da Silva, Anny Suellen Gomes.

Modelagem estatística para dados circulares / Anny Suellen Gomes da Silva. – 2019.
60 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Ciências, Curso de Estatística, Fortaleza, 2019.

Orientação: Prof. Dr. Gualberto Segundo Agamez Montalvo.

1. Dados circulares. 2. Estimadores de máxima verossimilhança.. I. Título.

CDD 519.5

ANNY SUELLEN GOMES DA SILVA

MODELAGEM ESTATÍSTICA PARA DADOS CIRCULARES

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Estatística do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Estatística.

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. Gualberto Segundo Agamez
Montalvo (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. José Roberto Silva dos Santos
Universidade Federal do Ceará (UFC)

Prof. Dr. Juvêncio Santos Nobre
Universidade Federal do Ceará (UFC)

À minha família, por sua capacidade de acreditar e investir em mim. Ao meu noivo, pela paciência e apoio. Aos meus professores e, finalmente, a minha amiga Rebeca que sempre me ajudou durante toda a graduação.

AGRADECIMENTOS

Primeiramente, ao Pai Celestial que sempre esteve e está ao meu lado em toda a minha vida, principalmente neste percurso acadêmico.

À minha família, por todo apoio, confiança e ensinamentos para ter coragem de prosseguir. São as pessoas que sempre acreditaram no meu potencial e me deram forças para nunca desistir dos meus sonhos.

Ao meu noivo, por seu amor, incentivo e compreensão em todos os momentos de estudo e preocupações.

Aos meus amigos Fábio, Lucas, Luan pela amizade e auxílio constante, mas principalmente a Rebeca, por sempre estar ao meu lado durante toda a graduação me ajudando em tudo que precisava. Não foi fácil enfrentar os desafios, mas nossa amizade recíproca e constante incentivo para continuar lutando é o que nos fez terminar juntas esse período acadêmico.

Aos professores Gualberto Segundo Agamez Montalvo e João Maurício Araujo Mota, pela dedicação, humildade e paciência durante a graduação, como também para a realização deste trabalho.

Aos professores José Roberto Silva dos Santos e Juvêncio Santos Nobre por aceitarem ao convite de avaliar este trabalho.

“Seu futuro é tão brilhante quanto sua fé.”

(Thomas S. Monson)

RESUMO

Em diversas áreas do conhecimento estão presentes dados angulares. A análise deste tipo de observação é distinta da que é usada para dados lineares, ou seja, com suporte em \mathbb{R}^2 . Neste estudo foram apresentados conceitos, representações gráficas, assim como algumas estatísticas resumo apropriadas para dados circulares. Além disso, o modelo probabilístico da distribuição Normal Circular foi estudado, visto que possui importantes propriedades teóricas e aplicações práticas. Finalmente, foi realizado um estudo de simulação com a finalidade de explorar o comportamento dos estimadores de máxima verossimilhança da Normal Circular.

Palavras-chave: Dados circulares. Normal Circular. Medidas resumo. Estimadores de máxima verossimilhança.

ABSTRACT

Angular data are common in many research fields. The analysis of this type of observation is different from that used for linear data, that is, with support in \mathbb{R}^2 . In the present work were presented concepts, graphical representations, as well as some summary statistics appropriate for directional data. In addition, the circular normal distribution was studied, since it has important theoretical properties and practical applications. Finally, a simulation study was carried out in order to explore the behavior of the maximum likelihood estimators of circular normal.

Keywords: Circular data. Circular normal. Measures summary. Maximum Likelihood Estimators.

LISTA DE FIGURAS

Figura 1 – Média aritmética dos ângulos 345° e 15°	15
Figura 2 – Relação entre coordenadas retangulares e polares.	17
Figura 3 – Pontos na circunferência de um círculo unitário para dados não agrupados.	17
Figura 4 – Gráfico de barras dos dados da Tabela 1	18
Figura 5 – Diagrama de rosas dos dados da Tabela 1	19
Figura 6 – Média aritmética dos ângulos 15° e 345°	19
Figura 7 – A média direcional \bar{x}_o e comprimento médio resultante \bar{R}	20
Figura 8 – Mediana para a mensuração dos nove ensaios da roleta.	27
Figura 9 – Representação do i -ésimo ponto P_i pertencente ao círculo unitário.	31
Figura 10 – Distribuição Normal Circular para $\mu = 0$ e diferentes κ	38
Figura 11 – Função acumulada da Normal Circular para diferentes κ	40
Figura 12 – Gráfico de barras para a localidade 1	45
Figura 13 – Gráfico de barras para a localidade 9.	46
Figura 14 – Gráfico de dispersão, diagrama de rosas e PP-plot para a localidade 1.	47
Figura 15 – Gráfico de dispersão, diagrama de rosas e PP-plot para a localidade 9.	48
Figura 16 – Gráfico de dispersão, diagrama de rosas e PP-plot para a variável lordose lombar.	50

LISTA DE TABELAS

Tabela 1 – O número de ocorrências de chuva de 1 segundo ou mais por hora nos U.S.A 1908-37 (Dyck e Mattice, 1941)	18
Tabela 2 – Direções de fuga de 15 pombos-correio, lançados a pouco mais de 16 quilômetros ao noroeste de seu pombal	23
Tabela 3 – Ângulos de fuga de 714 patos britânicos (adaptado de Matthews, 1961) . . .	25
Tabela 4 – Resultados obtidos das medidas da Tabela 3	26
Tabela 5 – Resultados das simulações para diferentes tamanhos amostrais	43
Tabela 6 – As seis primeiras observações, em graus, do conjunto de dados <i>fisherB13c</i> para as 14 localidades.	44
Tabela 7 – Estatística resumo usual para as localidades do conjunto de dados <i>fisherB13</i>	45
Tabela 8 – Sumário das estatísticas resumo para o conjunto de dados <i>fisherB13c</i> para as localidades 1 a 7.	46
Tabela 9 – Sumário das estatísticas resumo para o conjunto de dados <i>fisherB13c</i> para as localidades 8 a 14.	47
Tabela 10 – Teste Watson de normalidade circular para a localidade 1	48
Tabela 11 – Teste Watson de normalidade circular para a localidade 9	48
Tabela 12 – As cinco primeiras observações das variáveis descritas anteriormente para a classe AB do conjunto de dados.	49
Tabela 13 – Sumário das estatísticas resumo da variável lordose lombar.	49
Tabela 14 – Teste Watson de normalidade circular para a variável lordose lombar	50

LISTA DE SÍMBOLOS

$I_p(\kappa)$	Função Bessel modificada do primeiro tipo de ordem p
$F(\theta)$	Função distribuição
$f(\theta)$	Função densidade circular
$NC(\mu, \kappa)$	Distribuição Normal Circular com média direcional μ e parâmetro de concentração κ
m_p	p -ésimo momento trigonométrico
\bar{R}	Comprimento médio resultante
S	Soma dos senos dos ângulos
\bar{S}	Média amostral dos senos dos ângulos
C	Soma dos cossenos dos ângulos
\bar{C}	Média amostral dos cossenos dos ângulos
\bar{x}_o	Direção média
ξ	Direção mediana
n	Tamanho da amostra
$A_1(\kappa)$	Relação entre $I_1(\kappa)$ e $I_0(\kappa)$
θ	Variável aleatória circular
η_j	Ângulos em radianos
T_i	Amplitude circular
S_o	Variância circular
σ	Desvio padrão circular
φ_p	Função característica
$\overline{\varphi_p}$	Conjugado complexo de φ_p

SUMÁRIO

1	INTRODUÇÃO	14
2	DADOS CIRCULARES	16
2.1	Introdução	16
2.2	Representação gráfica	17
2.3	Medidas descritivas	19
2.3.1	<i>Medidas de tendência central</i>	20
2.3.1.1	<i>Média direcional</i>	20
2.3.1.2	<i>Dados não agrupados</i>	22
2.3.1.3	<i>Dados agrupados</i>	24
2.3.1.4	<i>Mediana</i>	27
2.3.1.5	<i>Moda</i>	28
2.3.2	<i>Medidas de dispersão</i>	29
2.3.2.1	<i>Amplitude Circular</i>	29
2.3.2.2	<i>Variância circular</i>	30
2.3.2.3	<i>Desvio padrão circular</i>	31
2.4	Distribuições Circulares	32
2.5	Função distribuição	32
2.6	Momentos trigonométricos	32
2.7	Função característica	36
3	DISTRIBUIÇÃO NORMAL CIRCULAR	37
3.1	Função Bessel	37
3.2	Função densidade de probabilidade	37
3.3	Função distribuição acumulada	39
3.4	Moda	40
3.5	Estimação de parâmetros da distribuição normal circular	41
3.5.1	<i>Simulação</i>	42
4	RESULTADOS	44
4.1	<i>fisherB13c</i>	44
4.2	UCI Machine Learning Repository - dados biomédicos	49
5	CONCLUSÕES E TRABALHOS FUTUROS	51

REFERÊNCIAS	52
APÊNDICES	54
APÊNDICE A – Códigos para os gráficos e medidas descritivas	54
APÊNDICE B – Códigos utilizados na aplicação	58
ANEXOS	60

1 INTRODUÇÃO

A nomenclatura *circular* para dados circulares ou direcionais é usada para que haja distinção dos dados lineares, os quais são dados com domínio em \mathbb{R}^2 . Medidas dadas em ângulos constituem um tipo de dados circulares, contudo estes dados não são necessariamente registrados inicialmente como ângulos. Dados circulares ou direcionais ocorrem de inúmeras formas, por exemplo, as direções do vento, orientações e migrações de pássaros e animais, os horários de chegada dos pacientes em um hospital, a incidência de uma doença ao longo do ano (em que o calendário é considerado relógio do ano), estudos de mapas mentais, entre outros. Além disso, a aplicação de dados circulares pode ser vista em diversos campos científicos, como em Demografia (COLEMAN; HASKEY, 1986), Ornitologia (BERGIN, 1991), Psicologia (MARDIA, 1972), Biologia (BATSCHELET, 1981), entre outras áreas.

Um dos trabalhos mais importantes sobre dados direcionais, a qual é amplamente utilizada para resolver os problemas aplicados a este tipo de observação foi introduzida por Von Mises em 1918, na qual foi proposta a chamada distribuição Normal Circular ou Normal Arqueada, sendo uma caracterização análoga a distribuição normal.

Dados circulares ou direcionais são representados como pontos na circunferência de um círculo unitário centrado na origem ou como vetores unitários conectando a origem a esses pontos, uma vez que uma direção não possui magnitude. Análogo a variável real, em que cada valor define um ponto na reta, cada ângulo define um ponto na circunferência do círculo unitário. Entretanto, devido à natureza periódica dos dados circulares, medidas com respeito à escolha de uma direção zero arbitrária e senso de rotação nos mostram a utilizar ferramentas e medidas estatísticas que preservem estas características. Dessa forma, diversos autores como Mardia (1972), N. I. Fisher (1993), Mardia e Jupp (1999) e Jammalamadaka e SenGupta (2001) fornecem métodos estatísticos para a análise de dados circulares.

A análise de dados circulares exige o uso de ferramentas e técnicas estatísticas específicas. Considere dois ângulos, medidos em um experimento 345° e 15° , de forma que deseja-se calcular a média destas observações. Utilizar os métodos usuais mais comuns pode torna-se inadequado, visto que uma média aritmética simples destes dados é 180° , sendo diametralmente oposta ao valor esperado. Dessa forma, torna-se uma medida equivocada.

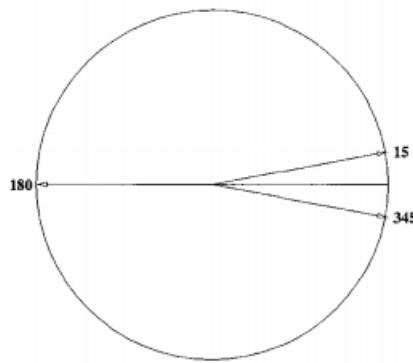


Figura 1 – Média aritmética dos ângulos 345° e 15°

Este trabalho tem como objetivo estudar e aplicar (manual e computacionalmente) as medidas apropriadas para os dados circulares, como também abordar os conceitos estatísticos da distribuição Normal Circular. Para ilustrar a teoria estudada, serão apresentados os resultados obtidos a partir das análises quantitativas e gráficas dos conjuntos de dados *fisherB13c* do pacote *circular* e os dados biomédicos encontrado na plataforma UCI Machine Learning Repository e alguns resultados obtidos em simulação (AGOSTINELLI; LUND, 2017; DHEERU; TANISKIDOU, 2017)

A monografia encontra-se estruturada da seguinte forma. No Capítulo 2, apresentamos os principais conceitos relacionados a dados circulares, como algumas representações gráficas e suas estatísticas resumo. No Capítulo 3, mostramos a distribuição Normal Circular, como também alguns conceitos da distribuição e avaliar o desempenho dos estimadores de máxima verossimilhança dos parâmetros. No Capítulo 4, fizemos duas aplicações usando as estatísticas resumo e a adequabilidade da distribuição Normal Circular para as variáveis circulares dos dois conjuntos de dados. Posteriormente, estão as conclusões e anexos, como também os apêndices que incluem os códigos para a realização do trabalho.

2 DADOS CIRCULARES

Neste capítulo apresentamos os principais conceitos relacionados a dados circulares, assim como algumas representações gráficas e estatística resumo para este tipo de dados. As estatísticas resumo serão exemplificadas manual e computacionalmente.

2.1 Introdução

Observações circulares podem ser encontradas em distintas áreas do conhecimento, como por exemplo, na Biologia, Economia, Geografia, entre outras. Algumas aplicações nessas áreas são:

1. A direção de uma abelha voando no campo coletando seu pólen;
2. A produção de carro em determinado mês;
3. Estudos de monitoramento usando latitudes e longitudes;
4. As direções do magnetismo remanescente;
5. O número de mortes devido a uma doença em cada mês durante anos.

Portanto, faz-se necessário o estudo de medidas que possam descrever corretamente tais fenômenos, dado que as medidas estatísticas usuais associadas a dados com domínio em \mathbb{R}^2 podem não são viáveis para dados circulares (MARDIA, 1972).

Um dado de uma variável circular é composto por um determinado ponto P que pertence à circunferência de um círculo unitário de um sistema de coordenadas cartesianas que é constituído por um vetor unitário em \mathbb{R}^2 . A mesma observação pode ser vista como um ângulo α formado pelo vetor \vec{OP} e o semi-eixo positivo \vec{Ox} no sentido anti-horário (Ver Figura 2). Portanto, o par ordenado (x,y) pode ser representado por $(r\cos(y), r\sin(y))$ em coordenadas polares.

Na análise de dados circulares ou direcionais é importante observarmos o sentido do ângulo e não a magnitude do vetor, uma vez que a mesma nos dá o comprimento do segmento de reta e o sentido nos dá o ângulo que a reta forma com o eixo x . Portanto, por conveniência, esses vetores podem ser considerados como unitários, isto é, o raio da circunferencia é igual a 1 ($r = 1$ na Figura 1) (FISHER, 1993).

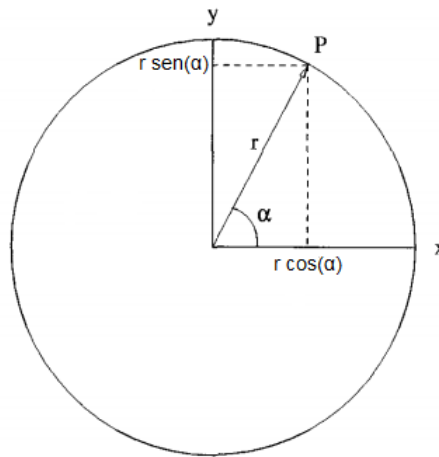


Figura 2 – Relação entre coordenadas retangulares e polares.

2.2 Representação gráfica

Os dados circulares são representados por gráficos e diagramas. Para dados agrupados a representação gráfica é dada por histogramas circulares ou diagrama de rosas. O histograma circular assemelha-se ao histograma usual, em que as barras representam as frequências de cada classe, entretanto, para dados direcionais as bases de cada uma destas barras estão sobre a circunferência de um círculo unitário.

O diagrama de rosas é um tipo de histograma circular formado por setores cujas áreas são proporcionais as frequências de cada classe e os raios são proporcionais à raiz quadrada de cada frequência (FISHER, 1993). Para dados não agrupados, as observações são representadas por pontos na circunferência de um círculo unitário, sendo atribuída a mesma massa para cada observação. Um exemplo é apresentado na Figura 3 (MARDIA, 1972).

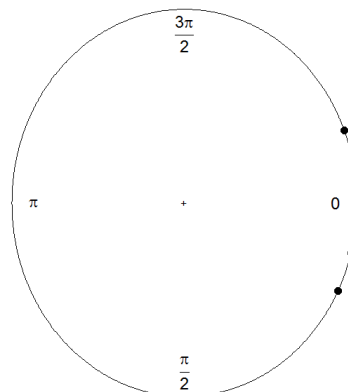


Figura 3 – Pontos na circunferência de um círculo unitário para dados não agrupados.

Exemplo 1 : Os dados da Tabela 1 representam os meses com seus respectivos ângulos e a frequência do número de ocorrências de chuva de 1 segundo ou mais por hora nos U.S.A 1908-37 (Dyck e Mattice, 1941). Cada dia do mês que corresponde aos 365 dias do ano equivale a 1°. Estes dados são utilizados para exemplificar a representação gráfica dos dados circulares.

Tabela 1 – O número de ocorrências de chuva de 1 segundo ou mais por hora nos U.S.A 1908-37 (Dyck e Mattice, 1941)

Meses	Ângulos	Número de Ocorrências
Janeiro	0° † 30°	101
Fevereiro	30° † 60°	104
Março	60° † 90°	231
Abril	90° † 120°	406
Mai	120° † 150°	683
Junho	150° † 180°	1225
Julho	180° † 210°	1475
Agosto	210° † 240°	1381
Setembro	240° † 270°	907
Outubro	270° † 300°	382
Novembro	300° † 330°	195
Dezembro	330° † 360°	145
Total		7235

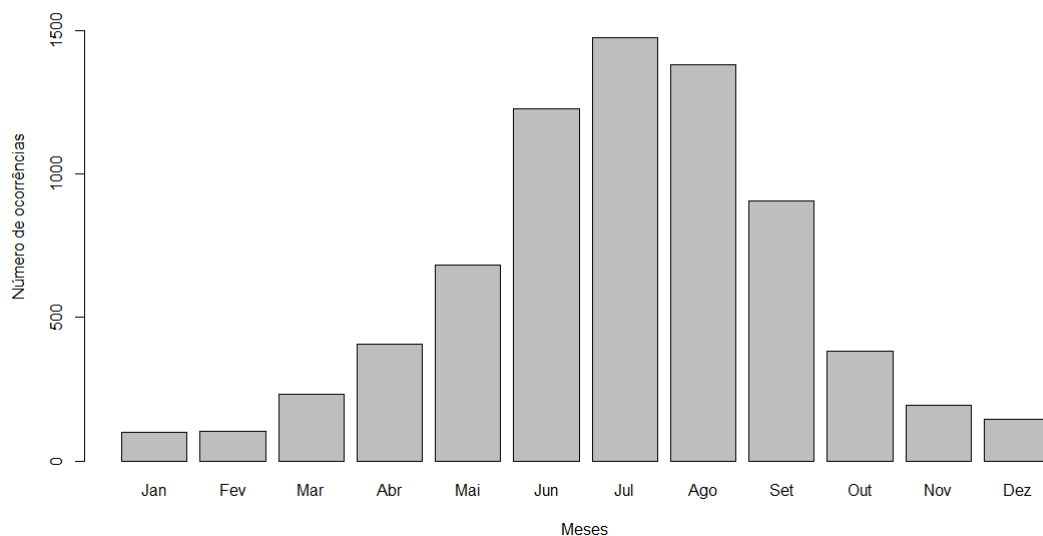


Figura 4 – Gráfico de barras dos dados da Tabela 1

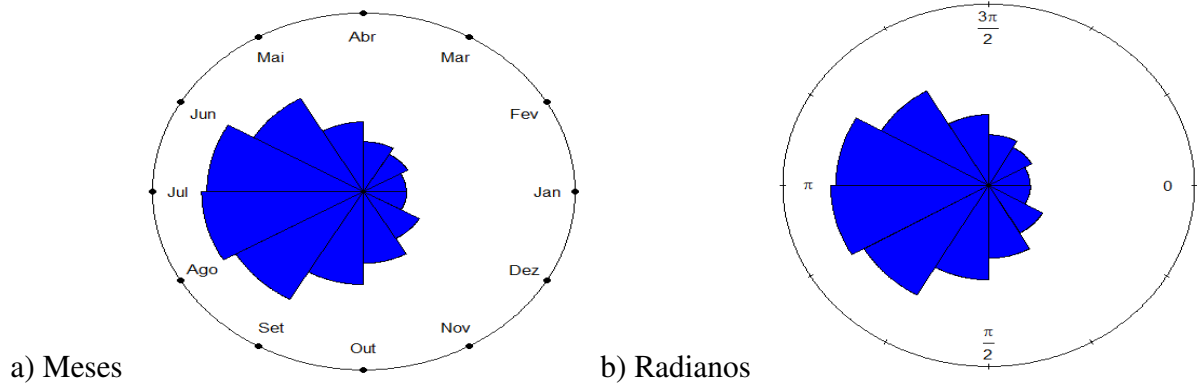


Figura 5 – Diagrama de rosas dos dados da Tabela 1

2.3 Medidas descritivas

Dados direcionais têm suas próprias medidas estatísticas. Entretanto, é comum que pensemos em usar as mesmas medidas utilizadas em dados lineares. No Exemplo 2 podemos perceber a inadequação de tais medidas para a análise de dados circulares.

Exemplo 2: Medido da maneira padrão, (com leste como direção zero e no sentido anti-horário como o sentido positivo de rotação), suponha que dois pássaros voaram a 15° e 345° . Calcular a média para estes dados (JAMMALAMADAKA; SENGUPTA, 2001).

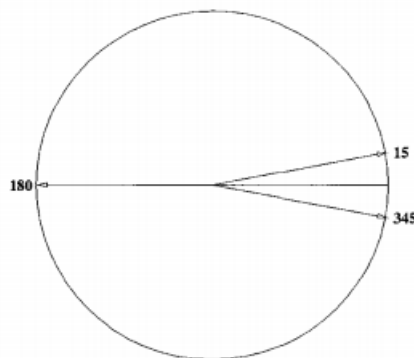


Figura 6 – Média aritmética dos ângulos 15° e 345°

Como podemos ver na Figura 6 a média aritmética destes dados, considerando que as duas observações apontam para o leste, é diametralmente oposta ao esperado, pois é direcionada ao ângulo de 180° . Dessa forma, a média aritmética pode não ser uma medida adequada para dados circulares.

2.3.1 Medidas de tendência central

Medidas de posição são valores que representam a tendência de aglomeração das observações ou dados. Algumas medidas de tendência central importantes são: média, mediana e moda. Vale ressaltar que, estas medidas são diferentes em comparação aos dados em \mathbb{R}^2 .

2.3.1.1 Média direcional

A partir do Exemplo 2, é perceptível que a média aritmética dos dados não é uma medida que indica o "centro" para um conjunto de observações direcionais. Uma medida apropriada e significativa da direção média para um conjunto de direções que são unimodais, é obtida tratando os dados como vetores unitários e usando a direção do seu vetor resultante. Seja θ_j o ângulo que pertence a um círculo unitário referente a um ponto P_j , para $j = 1, \dots, k$. A média direcional \bar{x}_o , é definida como a direção do vetor resultante $\overrightarrow{OP_1}, \dots, \overrightarrow{OP_k}$. As coordenadas cartesianas associadas a P_j são $(\cos(\theta_j), \text{sen}(\theta_j))$, $j = 1, \dots, k$. Dessa forma, o centro de gravidade desses pontos é (\bar{C}, \bar{S}) , em que,

$$\bar{C} = \sum_{j=1}^n \frac{\cos(\theta_j)}{n} \quad \text{e} \quad \bar{S} = \sum_{j=1}^n \frac{\text{sen}(\theta_j)}{n}.$$

Ao analisar a Figura 7 com \bar{x}_o definida, vemos que:

$$\cos(\bar{x}_o) = \frac{\bar{C}}{\bar{R}} \quad \text{e} \quad \text{sen}(\bar{x}_o) = \frac{\bar{S}}{\bar{R}}$$

em que, $\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2} = \frac{R}{n} \in [0, 1]$.

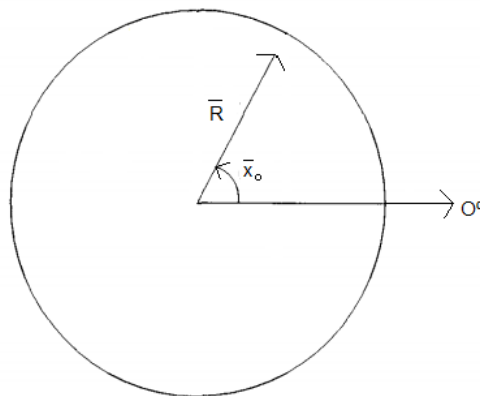


Figura 7 – A média direcional \bar{x}_o e comprimento médio resultante \bar{R}

Logo, \bar{x}_o é encontrado ao resolver a Equação 2.1 em função de \bar{S} e \bar{C} ,

$$\frac{\bar{S}}{\bar{C}} = \frac{\bar{R}\text{sen}(\bar{x}_o)}{\bar{R}\text{cos}(\bar{x}_o)} = \tan(\bar{x}_o) \quad (2.1)$$

ou seja, a média direcional é dada por:

$$\bar{x}_o = \begin{cases} \arctan\left(\frac{\bar{S}}{\bar{C}}\right), & \bar{S} \geq 0 \text{ e } \bar{C} > 0 \\ \frac{\pi}{2}, & \bar{C} = 0 \text{ e } \bar{S} > 0 \\ \arctan\left(\frac{\bar{S}}{\bar{C}}\right) + \pi, & \bar{C} < 0 \\ \arctan\left(\frac{\bar{S}}{\bar{C}}\right) + 2\pi, & \bar{S} < 0 \text{ e } \bar{C} \geq 0 \\ \text{Indefinido}, & \bar{C} = 0 \text{ e } \bar{S} = 0. \end{cases}$$

em que,

$$-\frac{\pi}{2} < \arctan\left(\frac{\bar{S}}{\bar{C}}\right) < \frac{\pi}{2}$$

A média direcional pode ser expressa em função de $C = \sum_{j=1}^n \cos(\theta_j)$ e $S = \sum_{j=1}^n \text{sen}(\theta_j)$, dessa forma sendo $R = \sqrt{C^2 + S^2}$. A vantagem de utilizar \bar{C} , \bar{S} e \bar{R} é que $\bar{R} \in [0, 1]$. No caso em que \bar{x}_o é indefinida, temos que $\bar{R} = 0$, ou seja, $\bar{C} = 0$ e $\bar{S} = 0$, isto nos mostra que as observações estão uniformemente distribuídas, dessa forma tem-se concentração mínima ocasionando variabilidade máxima.

Analogamente, note que \bar{x}_o como medida de tendência central possui propriedades importantes. Se os dados forem deslocados para uma certa quantidade, o valor de \bar{x}_o também muda na mesma quantidade. Para provar esta propriedade assume-se que $\theta' = \theta + \alpha$, em que $\alpha \in \mathbb{R}$, portanto temos que,

$$\text{sen}(\theta') = \text{sen}(\theta + \alpha) = \text{sen}(\theta)\cos(\alpha) + \text{sen}(\alpha)\cos(\theta)$$

$$\text{cos}(\theta') = \text{cos}(\theta + \alpha) = \text{cos}(\theta)\cos(\alpha) - \text{sen}(\theta)\text{sen}(\alpha)$$

consequentemente,

$$\begin{aligned}
\bar{C}' &= \sum_{j=1}^n \frac{\cos(\theta'_j)}{n} = \cos(\alpha) \sum_{j=1}^n \frac{\cos(\theta_j)}{n} - \sin(\alpha) \sum_{j=1}^n \frac{\sin(\theta_j)}{n} \\
&= \cos(\alpha) \bar{C} - \sin(\alpha) \bar{S} \\
&= \cos(\alpha) \cos(\bar{x}_o) \bar{R} - \sin(\alpha) \sin(\bar{x}_o) \bar{R} \\
&= \bar{R} \cos(\bar{x}_o + \alpha),
\end{aligned}$$

e

$$\begin{aligned}
\bar{S}' &= \sum_{j=1}^n \frac{\sin(\theta'_j)}{n} = \cos(\alpha) \sum_{j=1}^n \frac{\sin(\theta_j)}{n} + \sin(\alpha) \sum_{j=1}^n \frac{\cos(\theta_j)}{n} \\
&= \cos(\alpha) \bar{S} + \sin(\alpha) \bar{C} \\
&= \cos(\alpha) \sin(\bar{x}_o) \bar{R} + \sin(\alpha) \cos(\bar{x}_o) \bar{R} \\
&= \bar{R} \sin(\bar{x}_o + \alpha).
\end{aligned}$$

Assim,

$$\begin{aligned}
\bar{x}'_o &= \arctan\left(\frac{\bar{S}'}{\bar{C}'}\right) \\
&= \frac{\bar{R} \sin(\bar{x}_o + \alpha)}{\bar{R} \cos(\bar{x}_o + \alpha)} \\
&= \frac{\sin(\bar{x}_o + \alpha)}{\cos(\bar{x}_o + \alpha)} \\
&= \arctan(\tan(\bar{x}_o + \alpha)) \\
&= \bar{x}_o + \alpha,
\end{aligned}$$

em que, \bar{x}'_o é a média direcional dos dados deslocados.

2.3.1.2 Dados não agrupados

Exemplo 3: A Tabela 2 representa as direções de fuga de 15 pombos-correio, lançados a pouco mais de 16 quilômetros ao noroeste de seu pombal, em que η_j são os ângulos transformados para radianos.

Tabela 2 – Direções de fuga de 15 pombos-correio, lançados a pouco mais de 16 quilômetros ao noroeste de seu pombal

θ_j	η_j	$\cos(\eta_j)$	$\text{sen}(\eta_j)$
85°	1,4835	0,0872	0,9962
135°	2,3561	-0,7071	0,7071
135°	2,3561	-0,7071	0,7071
140°	2,4434	-0,7660	0,6428
145°	2,5307	-0,8192	0,5736
150°	2,6179	-0,8660	0,5000
150°	2,6179	-0,8660	0,5000
150°	2,6179	-0,8660	0,5000
160°	2,7925	-0,9397	0,3420
200°	3,4906	-0,9397	-0,3420
210°	3,6651	-0,8660	-0,5000
220°	3,8397	-0,7660	-0,6428
225°	3,9269	-0,7071	-0,7071
270°	4,7123	0,0000	-1,0000
285°	4,9741	0,2588	-0,9659
Total		-9,4700	1,3109

Usando os dados da Tabela 2 temos que $n = 15$, $\bar{C} = -9,4700 / 15 = -0,6313$ e $\bar{S} = 1,3109 / 15 = 0,0874$.

Além disso,

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2} = 0,6373,$$

$$\cos(\bar{x}_o) = \bar{C} / \bar{R} = -7,2237, \quad \text{e} \quad \text{sen}(\bar{x}_o) = \bar{S} / \bar{R} = 0,1371$$

Consequentemente, como $\bar{C} < 0$ e $\bar{S} < 0$, a média direcional é dada por:

$$\bar{x}_o = \arctan\left(\frac{\bar{S}}{\bar{C}}\right) + \pi = 172,1186^\circ$$

O pacote *circular* do software R (AGOSTINELLI; LUND, 2017) permite obter diferentes estatísticas resumo para dados circulares. A média direcional para dados não agrupados pode ser obtida utilizando a função *mean.circular* do dito pacote.

Média direcional da Tabela 2

```

> dados = circular(c(85,135,135,140,145,150,150,150,160,
200,210,220,225,270,285), units = "degrees")
> mean.circular(dados)
Circular Data:
Type = angles
Units = degrees
Template = none
Modulo = asis
Zero = 0
Rotation = counter
[1] 172.1186

```

2.3.1.3 Dados agrupados

A média direcional para dados agrupados é obtida usando o ponto médio de cada classe e sua respectiva frequência. Existem diferentes formas em que as observações podem ser expostas, por exemplo:

1. Quando o intervalo dos ângulos é $(0^\circ, 360^\circ)$;
2. Quando o intervalo é de $(0^\circ, 180^\circ)$;
3. Quando o período é expresso em tempo.

Para isso, as medidas \bar{C} e \bar{S} são multiplicadas pela frequência (f_j) do intervalo de classe. Entretanto, a média direcional dos dados é calculada da mesma forma que no caso dos dados não agrupados. Isto é,

$$\bar{C} = \frac{1}{n} \sum_{j=1}^n f_j \cos(\theta_j), \quad \bar{S} = \frac{1}{n} \sum_{j=1}^n f_j \sin(\theta_j) \quad \text{e} \quad \bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}.$$

Exemplo 4: A Tabela 3 corresponde aos ângulos de fuga de 714 patos britânicos (adaptado de Matthews, 1961). Os dados se enquadram no primeiro caso, ou seja, estão variando de $(0^\circ, 360^\circ)$.

Tabela 3 – Ângulos de fuga de 714 patos britânicos (adaptado de Matthews, 1961)

Direção		Frequência
0°	┆ 20°	40
20°	┆ 40°	22
40°	┆ 60°	20
60°	┆ 80°	9
80°	┆ 100°	6
100°	┆ 120°	3
120°	┆ 140°	3
140°	┆ 160°	1
160°	┆ 180°	6
180°	┆ 200°	3
200°	┆ 220°	11
220°	┆ 240°	22
240°	┆ 260°	24
260°	┆ 280°	58
280°	┆ 300°	136
300°	┆ 320°	138
320°	┆ 340°	143
340°	┆ 360°	69
Total		714

Na Tabela 4 apresentaremos os resultados obtidos das medidas necessárias para obtermos a média direcional para dados agrupados.

Usando os dados da Tabela 4, temos que $n = 714$ (soma das frequências), $\bar{C} = 356,8315/714 = 0,4997$ e $\bar{S} = -366,0377/714 = -0,5126$.

Além disso,

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2} = 0,7159,$$

$$\cos(\bar{x}_o) = \bar{C}/\bar{R} = 0,6981 \quad \text{e} \quad \sin(\bar{x}_o) = \bar{S}/\bar{R} = -0,7160$$

Conseqüentemente, como $\bar{C} > 0$ e $\bar{S} < 0$, a média direcional é dada por:

$$\bar{x}_o = \arctan\left(\frac{\bar{S}}{\bar{C}}\right) + 2\pi = 314,3^\circ$$

Tabela 4 – Resultados obtidos das medidas da Tabela 3

θ_j	η_j	f	$f \cos(\eta_j)$	$f \text{sen}(\eta_j)$
10	0,17	40	39,39	6,95
30	0,52	22	19,05	11,00
50	0,87	20	12,86	15,32
70	1,22	9	3,08	8,46
90	1,57	6	0,00	6,00
110	1,92	3	-1,03	2,82
130	2,27	3	-1,93	2,30
150	2,62	1	-0,87	0,50
170	2,97	6	-5,91	1,04
190	3,32	3	-2,95	-0,52
210	3,67	11	-9,53	-5,50
230	4,01	22	-14,14	-16,85
250	4,36	24	-8,21	-22,55
270	4,71	58	-0,00	-58,00
290	5,06	136	46,51	-127,80
310	5,41	138	88,70	-105,71
330	5,76	143	123,84	-71,50
350	6,11	69	67,95	-11,98
Total		714	356,8315	-366,0377

No software *R* a função *mean.circular* não está implementada para calcular a média direcional para dados agrupados. Portanto, foi implementada uma função para obtê-la. Os códigos são apresentados no Apêndice A.

Média direcional da Tabela 3

```
> media.circ(dados,f)
```

```
[1] 314.2704
```

É possível encontrar observações em que os ângulos variam somente no intervalo de $(0^\circ, 180^\circ)$, neste caso, temos que primeiramente dobrar os ângulos, ou seja, $\theta_i^* = 2\theta_i$. Posteriormente, encontram-se os respectivos pontos médios de cada intervalo de classe associados a θ_i^* (JAMMALAMADAKA; SENGUPTA, 2001). As medidas \bar{C}^* , \bar{S}^* e \bar{x}_o^* são quantidades relativas a θ_i^* que correspondem a \bar{C} , \bar{S} e \bar{x}_o . Portanto, a média direcional \bar{x}'_o para θ_i é obtida por,

$$\bar{x}'_o = \frac{\bar{x}_o^*}{2}$$

Há possibilidade que o pesquisador tenha interesse em expressar a média em termos de dias, por exemplo, a Tabela 1 que representa o número de ocorrências de chuva de 1 segundo ou mais por hora nos U.S.A 1908-37 nos 12 meses do ano. Dessa forma, é necessário que \bar{x}_o

seja expresso em dias da seguinte forma,

$$\bar{x}'_o = \frac{\bar{x}_o \times 365}{360}$$

Equivalentemente, \bar{x}_o pode ser convertida em horas e minutos ao utilizar o fato que 1° equivale a 4 minutos.

2.3.1.4 Mediana

A direção mediana ξ dos ângulos $\theta_1, \theta_2, \dots, \theta_n$ pode ser definida como um ângulo θ , tal que:

- Metade das observações esteja no arco $[\theta, \theta + \pi)$;
- A maioria das observações esteja mais perto de θ do que de $\theta + \pi$.

É importante avaliar a ordem em que os dados estão expostos, ou seja, é ajustada de maneira angular. Dessa forma, quando o tamanho amostral n é ímpar, a direção mediana é uma das observações, mais especificamente a $\frac{n+1}{2}$ -ésima observação, sendo a primeira, a mais distante de θ pertencente a qualquer um dos dois arcos ($[\theta, \theta + \pi)$ ou $[\theta + \pi, \theta + 2\pi)$). Quando n é par, é usual tomar a direção mediana como ponto médio das duas observações apropriadas (MARDIA, 1972).

Exemplo 5: Uma roleta foi autorizada a girar e suas posições de parada foram medidas em ângulos com direção fixa. a mensuração dos nove ensaios foram $43^\circ, 45^\circ, 52^\circ, 61^\circ, 75^\circ, 88^\circ, 88^\circ, 279^\circ, 357^\circ$.

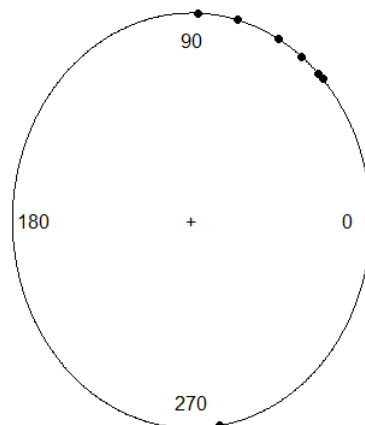


Figura 8 – Mediana para a mensuração dos nove ensaios da roleta.

Pode-se observar que o número de observações é ímpar, dessa forma é selecionado o ângulo que divide as observações (de maneira angular) ao meio, ou seja, a mediana deste

conjunto de dados é 52°.

A mediana circular para dados circulares pode ser obtida utilizando a função *median.circular*.

```

Mediana direcional dos nove ensaios da posição de parada de uma
roleta.
> dados = circular(c(43,45,52,61,75,88,88,279,357), units = "degrees")
> median.circular(dados)
Circular Data:
Type = angles
Units = degrees
Zero = 0
Rotation = counter
[1] 52
attr(,"medians")
[1] 52

```

2.3.1.5 Moda

A moda para dados direcionais agrupados é encontrada de forma semelhante aos dados lineares. A fórmula para encontrarmos a moda de um conjunto de dados é dada por

$$Moda = l + \frac{f_0 - f_{ant}}{2f_0 - f_{ant} - f_{post}} \times h,$$

em que, l é o limite inferior da classe modal, f_0 é a frequência da classe modal, f_{ant} e f_{post} são respectivamente as frequências das classes que antecedem e sucedem a classe modal, e h é a amplitude do intervalo de classe (MARDIA, 1972).

Para encontrar a moda da Tabela 3, o primeiro passo é encontrar a classe modal, ou seja, a classe (intervalo) que tem a maior frequência. Portanto, a maior concentração dos dados está no intervalo (320°,340°) com 143 observações. Dessa forma, a moda deste conjunto de dados é dada por,

$$Moda = 320 + \frac{143 - 138}{2 \times 143 - 138 - 69} \times 20,$$

dado que, $l = 320$, $f_0 = 143$, $f_{ant} = 138$, $f_{post} = 69$, e $h = 20$. Assim, a moda é 321,3°.

2.3.2 Medidas de dispersão

Uma maneira mais completa de apresentar os dados direcionais, além das medidas de tendência central, é apresentar também uma medida de dispersão. Estas medidas contêm informações que descrevem sobre até que ponto os valores da amostra estão agrupados em torno da média, entre outras finalidades. A seguir, introduziremos quantidades estatísticas tais como amplitude, variância e desvio circular (BATSCHELET, 1981).

2.3.2.1 Amplitude Circular

A amplitude circular é o comprimento do menor arco que contém todas as observações. Para determinarmos a distância circular inicialmente consideraremos os ângulos $\theta_1, \dots, \theta_j$ de forma ordenada $\theta(1) \leq \dots \leq \theta(n)$. Os comprimentos de arco entre observações adjacentes são:

$$T_i = \theta_{(i+1)} - \theta_{(i)}, \quad i = 1, \dots, n-1 \quad \text{e} \quad T_n = 2\pi - \theta_{(n)} + \theta_{(1)}.$$

Dessa forma, a amplitude circular é dada por:

$$H = 2\pi - \max(T_1, \dots, T_n)$$

Exemplo 7: A partir dos dados da Tabela 2 podemos encontrar a amplitude circular, dessa forma temos que,

$$\theta_i : 85^\circ, 135^\circ, 135^\circ, 140^\circ, 145^\circ, 150^\circ, 150^\circ, 150^\circ, 160^\circ, 200^\circ, 210^\circ, 220^\circ, 225^\circ, 270^\circ, 285^\circ$$

- $T_1 = 135^\circ - 85^\circ = 50^\circ$
- $T_2 = 135^\circ - 135^\circ = 0^\circ$
- $T_3 = 140^\circ - 135^\circ = 5^\circ$
- $T_4 = 145^\circ - 140^\circ = 5^\circ$
- $T_5 = 150^\circ - 145^\circ = 5^\circ$
- $T_6 = 150^\circ - 150^\circ = 0^\circ$
- $T_8 = 160^\circ - 150^\circ = 10^\circ$
- $T_9 = 200^\circ - 160^\circ = 40^\circ$
- $T_{10} = 210^\circ - 200^\circ = 10^\circ$
- $T_{11} = 220^\circ - 210^\circ = 10^\circ$

- $T_{12} = 225^\circ - 220^\circ = 5^\circ$
- $T_{13} = 270^\circ - 225^\circ = 45^\circ$
- $T_{14} = 285^\circ - 270^\circ = 15^\circ$
- $T_n = T_{15} = 360^\circ - 285^\circ + 85 = 160^\circ$

Portanto, temos que $H = 360^\circ - 160^\circ = 200^\circ$

A amplitude circular é obtida pela função *range.circular*.

Amplitude circular para a Tabela 2

```
x = circular(c(85,135,135,140,145,150,150,150,160,200,210,
220,225,270,285),units = "degrees")
> range.circular(x)
Circular Data:
Type = angles
Units = degrees
Zero = 0
Rotation = counter
[1] 200
```

2.3.2.2 Variância circular

Não só o sentido ao do vetor resultante R fornece uma direção média como acabamos de observar, mas seu comprimento. Seja P_i o ponto correspondente a θ_i no círculo unitário e α uma direção fixa. Vamos supor inicialmente que $\alpha = 0$ e suponha que P é o ponto correspondente no círculo. Uma medida da dispersão circular entre P_i e P é o menor dos dois ângulos que $\overrightarrow{OP_i}$ faz com \overrightarrow{OP} (MARDIA, 1972).

Ao analisar a Figura 9 vemos que,

$$\delta_i = \min(\theta_i, 2\pi - \theta_i) = \pi - |\pi - \theta_i|$$

Desde que $1 - \cos(\delta_i)$ é uma função monotonicamente crescente de δ_i temos que,

$$W = \frac{1}{n} \sum_{j=1}^n (1 - \cos \delta_j)$$

é considerada a medida de dispersão dos pontos P_i . Temos que ao mudarmos a direção zero,

$$W = \frac{1}{n} \sum_{j=1}^n (1 - \cos(\theta_j - \alpha))$$

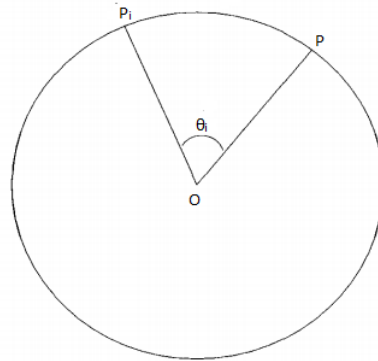


Figura 9 – Representação do i -ésimo ponto P_i pertencente ao círculo unitário.

Por propriedade de minimização a dispersão W é mínima quando $\alpha = \bar{x}_o$. Portanto podemos escrever o valor de W sob \bar{x}_o por S_o como,

$$S_o = 1 - \frac{1}{n} \sum_{j=1}^n (\cos(\theta_j - \bar{x}_o)) = 1 - \bar{R}$$

A demonstração deste resultado encontra-se no Apêndice A.

em que, S_o é a variância circular e \bar{R} é o comprimento médio resultante. Por fim,

$$0 \leq S_o \leq 1$$

O conceito é análogo ao da variância de dados lineares, no sentido de que quanto menor o valor da variância circular, maior a concentração nos dados.

A variância direcional pode ser obtida pela função *var.circular*.

Variância dos dados da Tabela 2

```
> dados = circular(c(85,135,135,140,145,150,150,150,160,
200,210,220,225,270,285), units = "degrees")
> var.circular(dados)
[1] 0.3626413
```

2.3.2.3 Desvio padrão circular

O desvio padrão circular amostral é uma transformação da variância S_o definida por:

$$\sigma = (-2 \ln(1 - S_o))^{\frac{1}{2}} = (-2 \ln(\bar{R}))^{\frac{1}{2}}$$

O desvio padrão direcional pode ser obtido pela função *sd.circular*.

Desvio padrão dos dados da Tabela 2

```
> dados = circular(c(85,135,135,140,145,150,150,150,160,
200,210,220,225,270,285), units = "degrees")
> sd.circular(dados)
[1] 0.9491287
```

2.4 Distribuições Circulares

As distribuições circulares teóricas são ditas absolutamente contínuas se existe uma função densidade de probabilidade, isto é, uma função contínua do ângulo θ que obedece as seguintes regras:

1. $f(\theta) \geq 0$
2. $\int_0^{2\pi} f(\theta)d\theta = 1$
3. $f(\theta) = f(\theta + 2\pi k)$, em que $k \in \mathbb{Z}$

2.5 Função distribuição

Seja W uma variável aleatória circular associada a um ângulo θ com suporte no intervalo $0 < \theta \leq 2\pi$. Podemos definir a função de distribuição F de W pela equação

$$F(\theta) = P(0 < W \leq \theta), \quad 0 < \theta \leq 2\pi \quad (2.2)$$

O ângulo θ está restrito ao intervalo $(0, 2\pi)$, entretanto podemos expandir o domínio para toda a reta, de modo que

$$F(\theta + 2\pi) - F(\theta) = 1 \quad \theta \in \mathbb{R} \quad (2.3)$$

Propriedades:

- $F(0) = 0$
- $F(2\pi) = 1$
- $\int_{\theta_1}^{\theta_2} f(\theta)d\theta = F(\theta_1) - F(\theta_2), \quad -\infty < \theta_1 \leq \theta_2 < \infty$

2.6 Momentos trigonométricos

Distribuições de probabilidade geralmente são descritas por seus momentos. O primeiro momento é dado por μ , ou seja, o valor esperado da variável aleatória integrável X , ou

seja, $E(X) \leq \infty$. Então temos que,

$$\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx,$$

$$\mu = \mathbb{E}(X^p) = \int_{-\infty}^{\infty} x^p f(x)dx$$

Os momentos trigonométricos utilizados nas variáveis com domínio em \mathbb{R}^2 são análogos aos que serão expostos para distribuições circulares, entretanto, momentos ordinários não podem ser aplicados devido à periodicidade envolvida na distribuição circular, portanto, um novo tipo de momentos é definido (MARDIA, 1972).

A fórmula de Euler mostra uma relação entre as funções trigonométricas seno e cosseno com a função exponencial, em que,

$$e^{itx} = \cos(tx) + i \operatorname{sen}(tx), \text{ em que } t \in \mathbb{R}$$

Portanto o primeiro momento trigonométrico circular amostral em torno da direção zero é dado da seguinte forma,

$$m_1 = \bar{C} + i\bar{S}.$$

Podemos reescrever o primeiro momento amostral em função do vetor resultante \bar{R} e a média direcional \bar{x}_o , em que

$$\begin{aligned} m_1 &= \bar{R} \cos(\bar{x}_o) + i\bar{R} \operatorname{sen}(\bar{x}_o) \\ &= \bar{R}(\cos(\bar{x}_o) + i \operatorname{sen}(\bar{x}_o)) \\ &= \bar{R} e^{i\bar{x}_o} \end{aligned}$$

Dessa forma, de maneira análoga, o momento trigonométrico circular populacional de ordem p em torno da direção zero é dado por,

$$m_p = a_p + ib_p$$

em que,

$$a_p = \sum_{j=1}^n \frac{\cos(p\theta_j)}{n} \quad \text{e} \quad b_p = \sum_{j=1}^n \frac{\operatorname{sen}(p\theta_j)}{n}.$$

Podemos definir o p-ésimo momento trigonométrico sobre a média direcional \bar{x}_o como,

$$m'_p = \bar{a}_p + i\bar{b}_p$$

em que,

$$\bar{a}_p = \sum_{j=1}^n \frac{\cos(p(\theta_j - \bar{x}_o))}{n} \quad \text{e} \quad \bar{b}_p = \sum_{j=1}^n \frac{\text{sen}(p(\theta_j - \bar{x}_o))}{n}.$$

O primeiro momento trigonométrico ($p = 1$), é obtido com a função *trigonometric.moment*.

Primeiro momento trigonométrico dos dados da Tabela 2

```
> dados = circular(c(85,135,135,140,145,150,150,150,160,
200,210,220,225,270,285), units = "degrees")
> trigonometric.moment(dados)
$mu
Circular Data:
Type = angles
Units = degrees
Template = none
Modulo = asis
Zero = 0
Rotation = counter
[1] 172.1186
$cos
[1] -0.6313382
$sin
[1] 0.08739681
$p
[1] 1
$n
[1] 15
```

A função *trigonometric.moment* retorna o p-ésimo momento trigonométrico amostral desejado adicionando o valor de p .

Segundo momento trigonométrico dos dados da Tabela 2

```
> dados = circular(c(85,135,135,140,145,150,150,150,160,
200,210,220,225,270,285), units = "degrees")
> trigonometric.moment(dados, p = 2)
$mu
Circular Data:
Type = angles
Units = degrees
Template = none
Modulo = asis
Zero = 0
Rotation = counter
[1] -71.07796
$cos
[1] 0.09137148
$sin
[1] -0.2665397
$p
[1] 2
$n
[1] 15
$call
trigonometric.moment(x = dados, p = 2)
```

2.7 Função característica

Seja Z uma variável aleatória com os valores em um círculo unitário, vamos identificar Z com uma variável aleatória θ , em que

$$Z = e^{i\theta} \quad 0 < \theta \leq 2\pi$$

Considere a seguinte função análoga a linha real,

$$\varphi(t) = \mathbb{E}(e^{it\theta}) = \int_0^{2\pi} e^{(it\theta)} f(\theta),$$

em que, $f(\theta)$ é a função distribuição de θ .

Pela propriedade da função distribuição dada em (2.2) podemos ver que,

$$\int_0^{2\pi} e^{(it(\theta+2\pi))} f(\theta) = \int_0^{2\pi} e^{(it\theta)} f(\theta),$$

sendo,

$$e^{2\pi it} = 1, \text{ em que } t \in \mathbb{Z}$$

Pela teoria das séries de Fourier para as funções periódicas podemos mostrar que é suficiente tomar t como um inteiro. Portanto, a função característica $\varphi(t)$ de θ é definida por,

$$\varphi_\theta(p) = \mathbb{E}(e^{ip\theta}) = \int_0^{2\pi} e^{(ip\theta)} f(\theta), \quad p = 0, \pm 1, \pm 2, \dots$$

Podemos reescrever a função característica da seguinte forma,

$$\varphi_p = \alpha_p + i\beta_p,$$

sendo,

$$\alpha_p = \mathbb{E}(\cos(p\theta)) = \int_0^{2\pi} \cos(p\theta) f(\theta)$$

$$\beta_p = \mathbb{E}(\sin(p\theta)) = \int_0^{2\pi} \sin(p\theta) f(\theta)$$

Assim, podemos notar as usuais propriedades de funções características. Para maior aprofundamento ver em (USHAKOV, 1999).

3 DISTRIBUIÇÃO NORMAL CIRCULAR

Um modelo estatístico para observações direcionais é a distribuição Normal Circular ou arqueada, a qual é uma distribuição de probabilidade contínua no círculo. Neste capítulo, apresentamos alguns conceitos da distribuição como, a função densidade, função distribuição acumulada, moda e os estimadores de máxima verossimilhança dos parâmetros.

3.1 Função Bessel

Uma família de funções que é de extrema importância neste estudo são as funções de Bessel modificada do primeiro tipo de ordem p , em que foi definida pela primeira vez por Daniel Bernoulli e generalizada por Friedrich Bessel e é dada por,

$$I_p(\kappa) = \frac{1}{\pi} \int_0^\pi \exp[\kappa \cos(x)] \cos(px) dx$$

Se $p = 0$, sabemos que $\cos(0x) = 1$, então obtêm-se a função $I_0(\kappa)$ (ROBERT, 1922).

3.2 Função densidade de probabilidade

A função densidade de probabilidade (fdp) de uma variável aleatória θ com distribuição normal circular é definida como:

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\theta - \mu)] \mathbf{I}_{[0, 2\pi]}(\theta), \quad (3.1)$$

em que, $\kappa (\kappa \geq 0)$ e $\mu \in [0, 2\pi)$. Dizemos então que θ , dados μ e κ , tem distribuição Normal Circular com ângulo médio μ e parâmetro de concentração κ , sendo denotada por $\theta \sim NC(\mu, \kappa)$. A Figura 10 apresenta a forma da densidade da distribuição Normal Circular com $\mu = 0$ e diferentes valores de κ . Podemos observar que quando $\kappa \rightarrow 0$, a distribuição Normal Circular converge para a distribuição circular uniforme no círculo $(0, 2\pi)$. Quando $\kappa \rightarrow \infty$, a distribuição converge para a distribuição degenerada no ponto $\theta = \mu$ (FISHER, 1993).

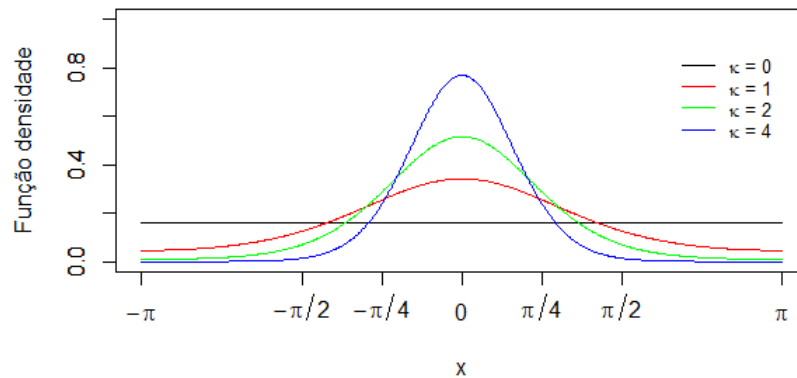


Figura 10 – Distribuição Normal Circular para $\mu = 0$ e diferentes κ .

As distribuições circulares obedecem as regras citadas na Seção 2.4, portanto devemos provar que a Função 3.1 é uma legítima função densidade de probabilidade.

1. $f(\theta) \geq 0$, para todo $\theta \in [0, 2\pi]$

Note que $\kappa > 0$ e $\mu \in [0, 2\pi)$, assim é verdade que $\exp[\kappa \cos(\theta - \mu)] > 0$, dado que $I_0(\kappa) > 0$. Portanto, implica que a densidade f é positiva.

2. $\int_0^{2\pi} \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\theta - \mu)] d\theta = 1$

Para avaliarmos ser uma legítima função densidade de probabilidade é necessário integrar a função dentro do suporte. Ao utilizarmos a transformação $v = \theta - \mu$, temos que:

$$\int_0^{2\pi} \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(v)] dv = \frac{1}{2\pi I_0(\kappa)} \int_{-\mu}^{2\pi-\mu} \exp[\kappa \cos(v)] dv$$

A função cosseno é periódica, assim podemos trocar o intervalo da integral que não altera o resultado. Logo, como já definida anteriormente, obtemos a função bessel modificada $I_0(\kappa)$.

$$\frac{1}{2\pi I_0(\kappa)} \int_0^{2\pi} \exp[\kappa \cos(v)] dv$$

Dessa forma,

$$\frac{I_0(\kappa)}{I_0(\kappa)} = 1$$

3. $f(\theta) = f(\theta + 2\pi k)$ para qualquer inteiro k

$$\begin{aligned} f(\theta) &= \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\theta - \mu)] \\ &= \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos((\theta + 2\pi) - \mu)] \end{aligned}$$

Seja,

- $\cos((\theta + 2\pi) - \mu) = \cos(\theta + 2\pi) \cos(\mu) + \sin(\theta + 2\pi) \sin(\mu)$
- $\cos(\theta + 2\pi) = (\cos(\theta) \cos(2\pi) - \sin(\theta) \sin(2\pi)) \cos(\mu)$
- $\sin(\theta + 2\pi) = (\sin(\theta) \cos(2\pi) + \cos(\theta) \sin(2\pi)) \sin(\mu)$

em que,

- $\sin(2\pi) = 0$
- $\cos(2\pi) = 1$

portanto,

$$\begin{aligned} \cos((\theta + 2\pi) - \mu) &= \cos(\theta) \cos(\mu) + \sin(\theta) \sin(\mu) \\ &= \cos(\theta - \mu) \end{aligned}$$

3.3 Função distribuição acumulada

A função distribuição acumulada da Normal Circular é dada por:

$$F(\theta) = P(\Theta \leq \theta) = \frac{\int_0^\theta \exp[\kappa \cos(\theta - \mu)]}{2\pi I_0(\kappa)} d\theta,$$

em que $0 < \theta < 2\pi$. A Figura 11 apresenta a forma da função distribuição acumulada da Normal Circular.

A função densidade de probabilidade pode ser expressa como uma série de funções de Bessel dada por:

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} \left(I_0(\kappa) + 2 \sum_{p=1}^{\infty} I_p(\kappa) \cos[p(\theta - \mu)] \right). \quad (3.2)$$

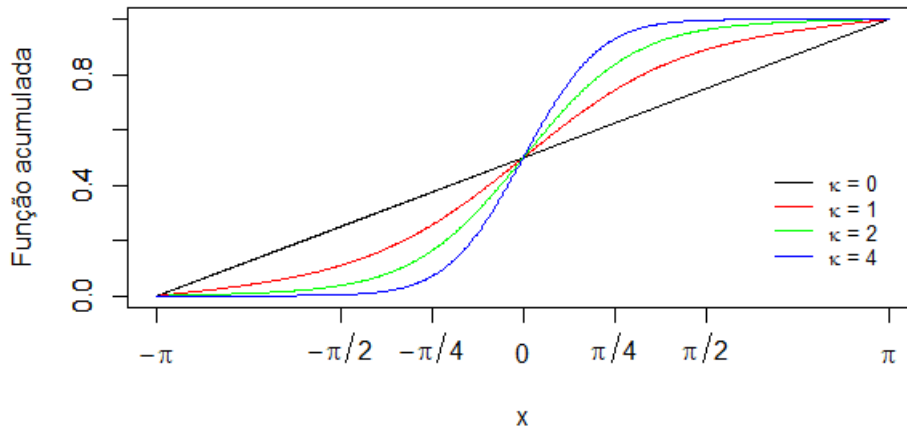


Figura 11 – Função acumulada da Normal Circular para diferentes κ .

Dessa forma, a função de distribuição acumulada da normal circular é obtida pela integração da Equação 3.2:

$$\begin{aligned}
 F(\theta) &= \int_0^x \frac{1}{2\pi I_0(\kappa)} \left(I_0(\kappa) + 2 \sum_{p=1}^{\infty} I_p(\kappa) \cos[p(\theta - \mu)] \right) d\theta \\
 &= \frac{1}{2\pi I_0(\kappa)} \left(\int_0^x I_0(\kappa) + 2 \sum_{p=1}^{\infty} \int_0^x I_p(\kappa) \cos[p(\theta - \mu)] \right) \\
 &= \frac{1}{2\pi I_0(\kappa)} \left(x I_0(\kappa) + 2 \sum_{p=1}^{\infty} \frac{I_p(\kappa) \text{sen}[p(\theta - \mu)]}{p} \right).
 \end{aligned}$$

3.4 Moda

Seja θ uma variável aleatória contínua com distribuição normal circular, a moda é dada pelo ponto no qual a função de densidade de probabilidade dada em 3.1 é máxima, dessa forma,

$$f'(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\theta - \mu)] [-\kappa \text{sen}(\theta - \mu)],$$

para $f'(\theta) = 0$

$$\exp[\kappa \cos(\theta - \mu)] [-\kappa \text{sen}(\theta - \mu)] = 0.$$

Quando $\theta = \mu$ tem-se que μ é o ponto crítico da função. Se a segunda derivada calculada no ponto crítico for negativa tem-se o ponto de máximo, portanto, o próximo passo é analisar a segunda derivada da função para verificar o ponto crítico encontrado,

$$f''(\theta) = -\frac{\kappa(-\kappa \exp[\kappa \cos(\theta - \mu)] \operatorname{sen}^2(\theta - \mu) + \exp[\kappa \cos(\theta - \mu)] \cos(\theta - \mu))}{2\pi I_o(\kappa)},$$

que avaliada em $\theta = \mu$, obtêm-se

$$f''(\theta) = \frac{-\kappa \exp[\kappa]}{2\pi I_o(\kappa)} < 0$$

Dessa forma, como a $f''(\theta)$ é menor que zero, então temos um ponto de máximo que é a moda da distribuição.

3.5 Estimação de parâmetros da distribuição normal circular

Seja a função densidade de probabilidade dada em 3.1, baseada em uma amostra independente e identicamente distribuída, neste estudo iremos usar o método de máximo verossimilhança para estimar os parâmetros da distribuição, entretanto existem estudos utilizando outros métodos, como por exemplo o Método dos Momentos (SPURR; KOUTBEIY, 1991), Estimação Bayesiana (MURALIDHARAN; PARIKH, 2007), entre outros (ABEYASEKERA; COLLETT, 1982). Portanto, a função de verossimilhança é dada por,

$$\begin{aligned} L(\mu; \theta_1, \dots, \theta_n) &= \prod_{i=1}^n \frac{1}{2\pi I_o(\kappa)} \exp[\kappa \cos(\theta_i - \mu)] \\ &= \left(\frac{1}{2\pi I_o(\kappa)} \right)^n \exp \left[\sum_{i=1}^n \kappa \cos(\theta_i - \mu) \right], \end{aligned}$$

em que $\theta_1, \dots, \theta_n$ é uma amostra aleatória simples de uma distribuição Normal Circular com parâmetros μ e κ . Para encontrar o estimador de máxima verossimilhança para μ , devemos encontrar o valor de μ para o qual a função de verossimilhança é máxima. Aplicando a função logaritmo natural (\ln) na função de verossimilhança $L(p; \theta_1, \dots, \theta_n)$, temos que:

$$\ln L(p, \theta_1, \dots, \theta_n) = -n \ln(2\pi) - n \ln(I_o(\kappa)) + \kappa \sum_{i=1}^n \cos(\theta_i - \mu).$$

Seja \bar{x}_o a média direcional dos dados temos que,

$$\sum_{i=1}^n \cos(\theta_i - \mu) = n\bar{R} \cos(\bar{x}_o - \mu)$$

Logo,

$$\ln L(p, \theta_1, \dots, \theta_n) = -n \ln(2\pi) - n \ln(I_o(\kappa)) + \kappa n \bar{R} \cos(\bar{x}_o - \mu)$$

Observa-se que quando $\mu = \bar{x}_o$ a função tem seu ponto de máximo, pois $\cos(0) = 1$. Assim podemos afirmar que \bar{x}_o é o estimador de máxima verossimilhança para μ .

Para a estimação do parâmetro κ , temos:

$$\frac{\partial \ln L(\kappa; \theta_1, \dots, \theta_n)}{\partial \kappa} = \frac{-n}{I_o(\kappa)} I'_o(\kappa) + n\bar{R} \cos(\bar{x}_o - \mu)$$

Igualando a zero para encontrar o estimador que maximiza a função:

$$\frac{\partial \ln L(\kappa; \theta_1, \dots, \theta_n)}{\partial \kappa} = 0$$

Então,

$$\frac{-n}{I_o(\hat{\kappa})} I'_o(\hat{\kappa}) + n\bar{R} \cos(\bar{x}_o - \hat{\mu}) = 0$$

Como $I'_o(\kappa) = I_1(\kappa)$, então: $\frac{I_1(\kappa)}{I_o(\kappa)} = A_1(\kappa)$. Ao substituir obtemos:

$$\frac{\partial \ln L(\kappa; \theta_1, \dots, \theta_n)}{\partial \kappa} = -nA_1(\kappa) + n\bar{R} \cos(\bar{x}_o - \hat{\mu})$$

E assim podemos encontrar o estimador de κ ,

$$\begin{aligned} -nA_1(\hat{\kappa}) + n\bar{R} \cos(\bar{x}_o - \hat{\mu}) &= 0 \\ &= -nA_1(\hat{\kappa}) + n\bar{R} \\ &= n(-A_1(\hat{\kappa}) + \bar{R}) \\ &= \bar{R} = A_1(\hat{\kappa}) \end{aligned}$$

Ou seja, o estimador de κ é dado por:

$$\hat{\kappa} = A_1^{-1}(\bar{R})$$

3.5.1 Simulação

Os resultados da Tabela 5 são referentes a 1000 simulações de amostras com distribuição normal circular de média $\mu = 0,2$ e $\kappa = 4$ para analisar o comportamento dos estimadores e máxima verossimilhança de μ e κ . No Apêndice A podemos entender como foi feita a simulação da distribuição.

Tabela 5 – Resultados das simulações para diferentes tamanhos amostrais

Resultados	Tamanho amostral (n)						
	10	20	50	100	200	500	1000
$E(\hat{\mu})$	0,9658	0,4701	0,2007	0,1994	0,2011	0,1977	0,1990
$E(\hat{\mu} - \mu)$	0,7658	0,2701	0,0007	-0,0005	0,0011	-0,0022	-0,0009
$E(\hat{\mu} - \mu)^2$	4,4014	1,5374	0,0055	0,0029	0,0015	0,0005	0,0003
$E(\hat{\kappa})$	5,6452	4,7537	4,2687	4,2081	4,1272	4,0958	4,0739
$E(\hat{\kappa} - \kappa)$	1,6452	0,7537	0,2687	0,2081	0,1272	0,0958	0,0739
$E(\hat{\kappa} - \kappa)^2$	12,5184	3,4529	0,6980	0,3721	0,1576	0,0603	0,0300

Neste estudo de simulação, podemos observar que os estimadores de máxima verossimilhança dos parâmetros da distribuição Normal Circular mostram um desempenho ruim para pequenas amostras. Entretanto, aumentando o tamanho amostral podemos observar que o estimador torna-se cada vez mais próximo do valor real do parâmetro. Assumindo-se que a função de verossimilhança satisfaz algumas propriedades matemáticas básicas, a partir dos resultados obtidos podemos classificar os estimadores como consistentes. Outros estudos de simulação foram desenvolvidos com diferentes valores dos parâmetros para o ângulo médio e a concentração. Nestes estudos os resultados foram equivalentes.

4 RESULTADOS

Uma análise descritiva para os dados direcionais utilizando o software R foi aplicada ao conjunto de dados *fisherB13* do pacote *circular*. Estas observações são provenientes de orientações dos cupinzeiros de *Amitermes Laurensis* em 14 locais em Cape York Peninsula, North Queensland. Além disso, a análise descritiva também foi aplicada ao conjunto de dados biomédicos construído pelo Dr. Henrique da Mota, médico especialista em coluna vertebral, encontrado no UCI Machine Learning Repository. O objetivo é estudar e analisar as estatísticas resumo da variável circular para cada localidade individualmente, como também as representações gráficas e adequabilidade dos dados a distribuição Normal Circular.

4.1 *fisherB13c*

A Tabela 6 mostra as seis primeiras orientações dos cupinzeiros de *Amitermes Laurensis* em 14 locais em Cape York Peninsula, North Queensland.

Tabela 6 – As seis primeiras observações, em graus, do conjunto de dados *fisherB13c* para as 14 localidades.

	Local 1	Local 2	Local 3	...	Local 13	Local 14
1	161	161	194	...	196	180
2	182	182	188	...	196	196
3	179	179	166	...	180	192
4	193	193	185	...	182	190
5	164	164	191	...	180	162
6	166	166	184	...	186	182

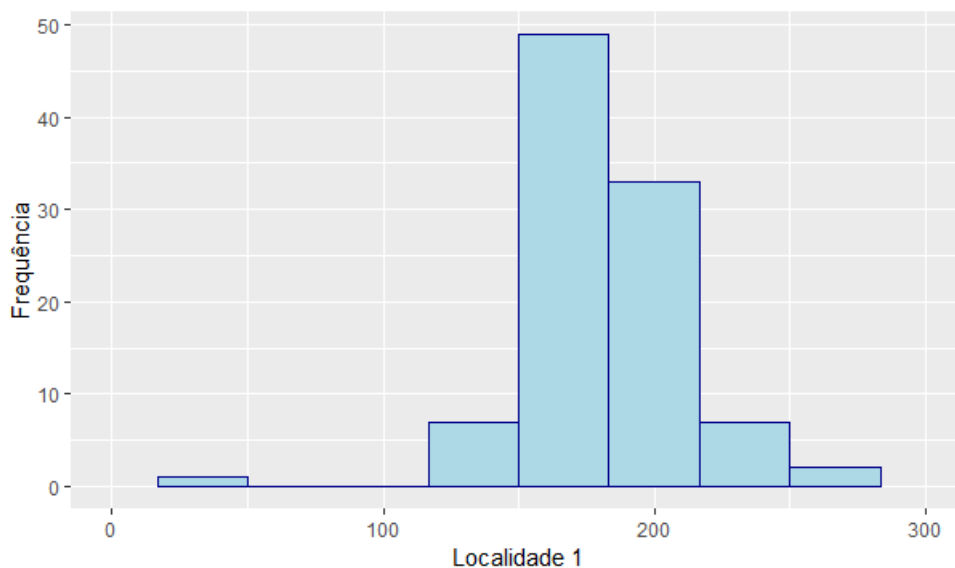
A Tabela 7 apresenta as estatísticas resumo usuais para as 14 localidades do banco de dados. A partir dos resultados obtidos das medidas de posição e dispersão das 14 localidades podemos observar que a localidade 10 é a mais homogênea em relação a variância e desvio padrão. Já a localidade 1 possui alta variabilidade comparada com as demais.

Tabela 7 – Estatística resumo usual para as localidades do conjunto de dados *fisherB13*.

	n	Média	Mediana	Variância	Desvio padrão
Local 1	100	182,82	180	1027,28	32,05
Local 2	50	171,08	173	275,14	16,59
Local 3	50	174,64	173,50	198,56	14,09
Local 4	50	171,58	171	201,68	14,20
Local 5	50	172,82	173	237,17	15,40
Local 6	50	177,82	177,50	134,56	11,60
Local 7	66	173,76	173,50	300,22	17,33
Local 8	48	174,44	177	412,08	20,30
Local 9	100	173,45	173	280,35	16,74
Local 10	50	172,50	172	85,77	9,26
Local 11	37	184,42	182,50	99,08	9,95
Local 12	31	176,19	178	293,09	17,12
Local 13	132	182,05	181	137,17	11,71
Local 14	92	179,36	178,50	167,22	12,93

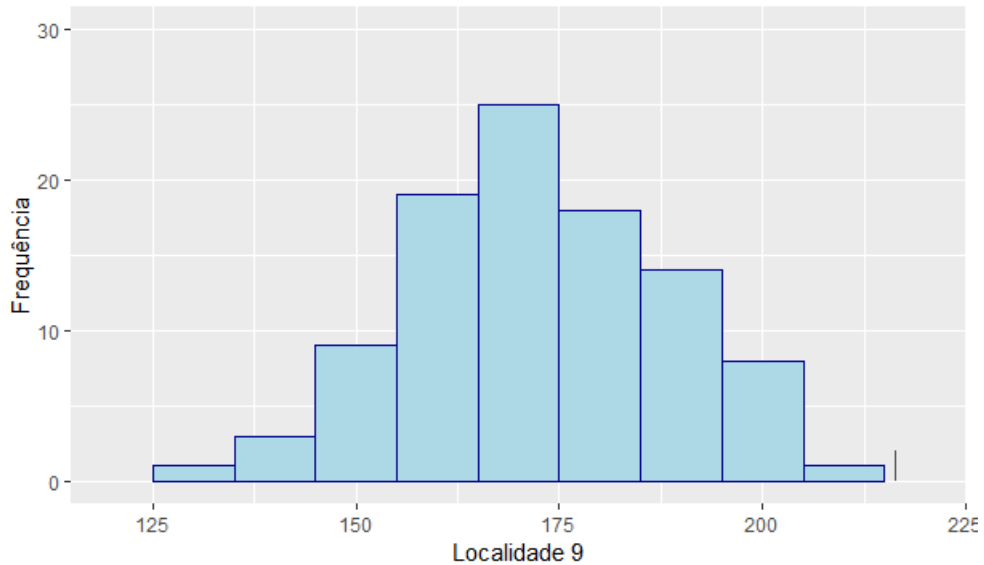
A Figura 12 representa o gráfico de barras para a localidade 1. A frequência decresce rapidamente em um dos lados, produzindo assim, uma calda mais longa. Podemos ver assimetria positiva dos dados. Uma medida de dispersão importante a ser analisada para esta localidade é a variância, visto que a quantificação indica que os valores estão distantes da média, assim não há indícios de normalidade dos dados.

Figura 12 – Gráfico de barras para a localidade 1



A Figura 13 representa o gráfico de barras para a localidade 9. Uma característica importante para a análise desta localidade é a proximidade entre a moda e a mediana dos dados. Há maior concentração dos dados a partir do centro, o que indica simetria. A representação gráfica e estatística resumo usual dos dados da localidade 9 indica que temos indícios de normalidade.

Figura 13 – Gráfico de barras para a localidade 9.



Entretanto, faz-se necessário analisar os dados de natureza angular, assim, para as análises posteriores utilizaremos o conjunto de dados *fisherB13c*. Analogamente, verificaremos a adequabilidade dos dados a distribuição Normal Circular.

As estatísticas resumo são calculadas por meio das funções dispostas no pacote *circular* e fornecidas nas Tabelas 8 e 9. A interpretação das estatísticas resumo para dados circulares é análoga as lineares.

Tabela 8 – Sumário das estatísticas resumo para o conjunto de dados *fisherB13c* para as localidades 1 a 7.

	Local 1	Local 2	Local 3	Local 4	Local 5	Local 6	Local 7
n	100	50	50	50	50	50	66
\bar{R}	0,14	0,03	0,07	0,09	0,04	0,04	0,18
\bar{x}_o	-177,20	171,22	174,66	171,50	172,86	177,80	174,07
T	242	81	65	59	82	54	94
S_o	0,12	0,04	0,03	0,03	0,03	0,02	0,04
σ	0,50	0,29	0,24	0,25	0,27	0,20	0,30
m_2	1,29	-16,75	-10,56	-17,46	-14,00	-4,54	-10,14

Tabela 9 – Sumário das estatísticas resumo para o conjunto de dados *fisherB13c* para as localidades 8 a 14.

	Local 8	Local 9	Local 10	Local 11	Local 12	Local 13	Local 14
n	48	100	50	37	31	132	92
\bar{R}	0,03	0,08	0,06	0,04	0,05	0,04	0,08
\bar{x}_o	174,33	173,40	172,50	-175,63	176,30	-178,01	179,39
T	137	82	55	50	104	88	83
S_o	0,06	0,04	0,01	0,01	0,04	0,02	0,02
σ	0,34	0,29	0,16	0,17	0,29	0,20	0,22
m_2	-11,50	-13,55	-15,02	8,43	-6,71	3,66	-1,06

Utilizando a função *pp.plot*, obtemos a distribuição empírica de um conjunto de dados contra a função de distribuição da Normal Circular. Portanto, podemos avaliar a adequabilidade dos dados a distribuição citada, ou seja, quanto mais próximo da reta, poderemos supor normalidade circular. A Figura 14 representa o gráfico de dispersão, diagrama de rosas e PP-plot para a localidade 1, podemos notar que os dados não apresentam normalidade circular. Entretanto, na Figura 15 que apresentam as mesmas representações gráficas para a localidade 9, podemos supor normalidade circular para os dados.

Figura 14 – Gráfico de dispersão, diagrama de rosas e PP-plot para a localidade 1.

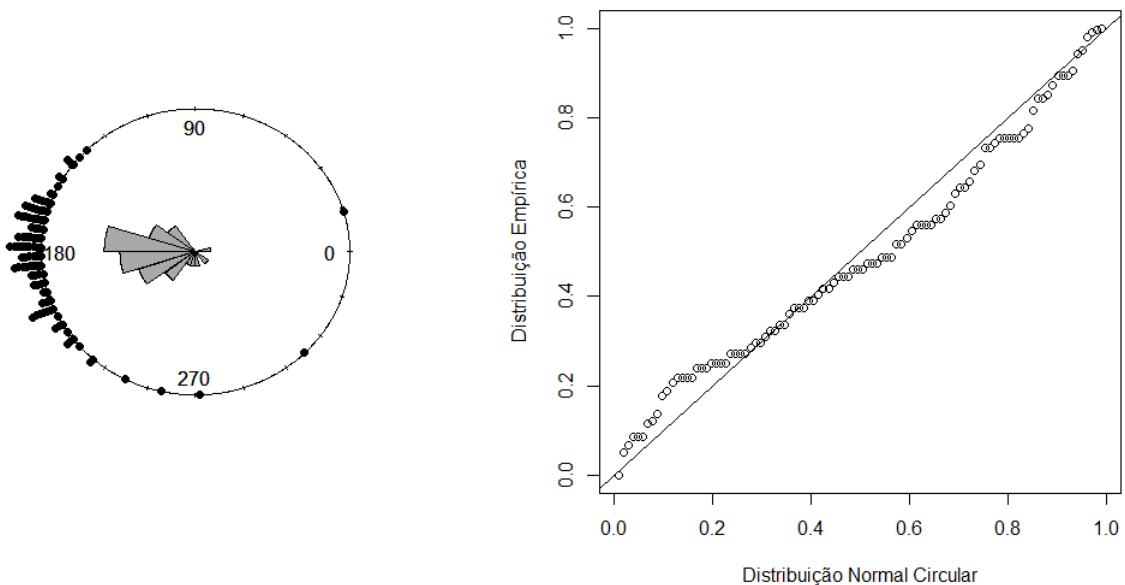
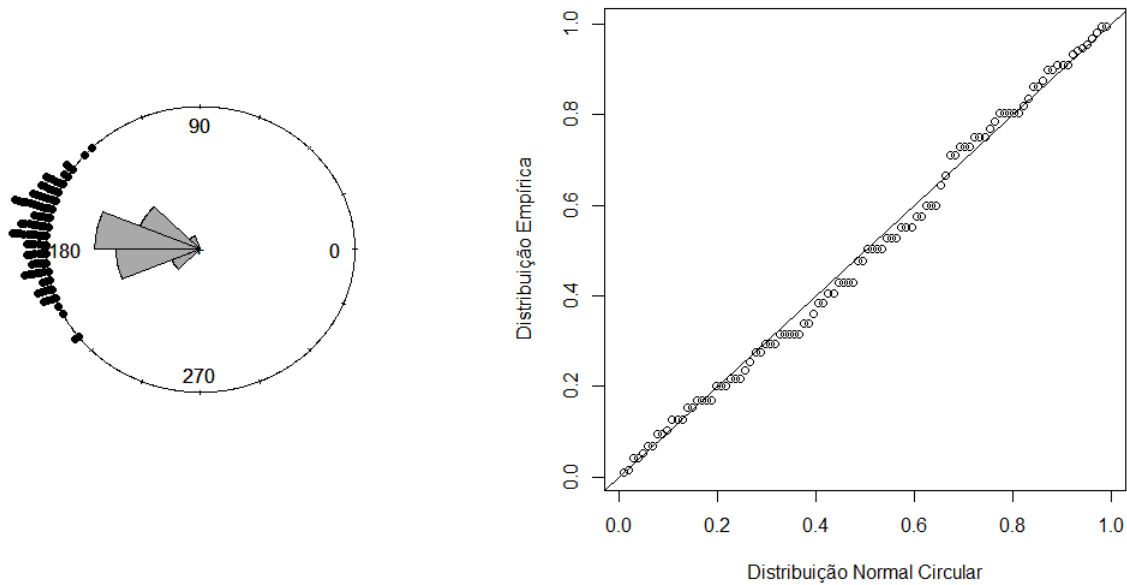


Figura 15 – Gráfico de dispersão, diagrama de rosas e PP-plot para a localidade 9.



Adicionalmente, foi feito o teste de normalidade circular (Watson's Test) para verificar o inferido na parte gráfica. Podemos observar na Tabela 10 que a estatística do teste foi maior que o valor crítico, dessa forma rejeita-se a hipótese de normalidade circular. Para a localidade 9, como visto na Tabela 11, o teste não rejeitou a hipótese de normalidade circular, sendo o valor crítico menor que a estatística. Para mais esclarecimentos sobre o teste, ver (JAMMALAMADAKA; SENGUPTA, 2001), por exemplo,

Tabela 10 – Teste Watson de normalidade circular para a localidade 1

Teste Watson para a distribuição Normal Circular	
Estatística do teste	0,2214
Valor crítico ¹	0,113

Tabela 11 – Teste Watson de normalidade circular para a localidade 9

Teste Watson para a distribuição Normal Circular	
Estatística do teste	0,0428
Valor crítico ²	0,117

¹ Nível de significância: 0,05

² Nível de significância: 0,05

4.2 UCI Machine Learning Repository - dados biomédicos

O conjunto de dados biomédicos construído pelo Dr. Henrique da Mota, médico especialista em coluna vertebral, possui seis atributos biomecânicos derivados da forma e orientação da pelve e coluna lombar para cada paciente, são estes nesta ordem, incidência pélvica (X_1), inclinação pélvica (X_2), ângulo da lordose lombar (X_3), inclinação sacral (X_4), raio pélvico (X_5) e grau de espondilolistese (X_6). A seguinte convenção é usada para os rótulos de classe: DH (Hérnia de Disco), Espondilolistese (SL), Normal (NO) e Anormal (AB). Neste estudo, o atributo biomecânico analisado é o ângulo de lordose lombar para a classe AB. A Tabela 12 apresenta as cinco primeiras observações dos seis atributos biomecânicos da classe AB.

Tabela 12 – As cinco primeiras observações das variáveis descritas anteriormente para a classe AB do conjunto de dados.

X_1	X_2	X_3	X_4	X_5	X_6	X_7
63	22,6	39,6	40,5	98,7	-0,25	AB
39,1	10,1	25	29	114	4,56	AB
68,8	22,2	50,1	46,6	106	-3,53	AB
69,3	24,6	44,3	44,6	102	11,2	AB
49,7	9,65	28,3	40,1	108	7,92	AB

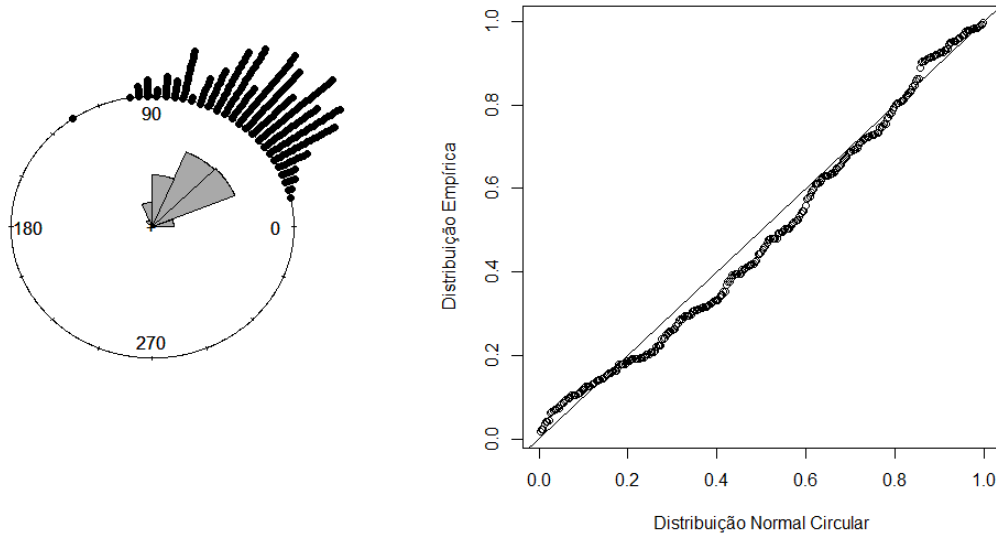
A Tabela 13 apresenta as estatísticas resumo para a variável lordose lombar. A partir destas medidas podemos observar um certo grau de variabilidade dos dados. Além disso, a média e mediana direcional não estão tão próximas, o que pode indicar não normalidade dos dados.

Tabela 13 – Sumário das estatísticas resumo da variável lordose lombar.

Lordose lombar	
n	310
\bar{x}_o	51,74
ξ	49,56
T	111,74
S_o	0,05
σ	0,32

A Figura 16 representa gráfico de dispersão, diagrama de rosas e PP-plot para a variável lordose lombar. Podemos notar que os dados não apresentam normalidade circular ao compararmos a distribuição empírica dos dados e o melhor ajuste da distribuição.

Figura 16 – Gráfico de dispersão, diagrama de rosas e PP-plot para a variável lordose lombar.



Adicionalmente, foi feito o teste de normalidade circular para verificar o inferido na parte gráfica da variável. Dessa forma, podemos observar na Tabela 14 que a estatística do teste foi maior que o valor crítico, dessa forma rejeita-se a hipótese de normalidade circular para a variável lordose lombar.

Tabela 14 – Teste Watson de normalidade circular para a variável lordose lombar

Teste Watson para a distribuição Normal Circular	
Estatística do teste	0,2782
Valor crítico ³	0,117

³ Nível de significância: 0,05

5 CONCLUSÕES E TRABALHOS FUTUROS

Pode-se notar que, de fato, é necessário o uso de uma análise estatística adequada para dados circulares, bem como novos modelos probabilísticos. Além disso, é imprescindível evitar possíveis erros de análise em diversas áreas do conhecimento.

O estudo de simulação para os estimadores de máxima verossimilhança dos parâmetros da distribuição Normal Circular nos mostram as similaridades com a distribuição Normal na reta. Além disso, a interpretação das estatísticas resumo e das representações gráficas para dados circulares são similares aos dados lineares, o que nos possibilita maior facilidade.

Neste trabalho, a adequabilidade da distribuição Normal Circular para o conjunto de dados do pacote *circular* realizada através das análises descritivas e utilizando a representação gráfica do PP-plot foi razoável para algumas das localidades, entretanto, para o conjunto de dados biomédicos a distribuição não foi adequada. Dessa forma, faz-se necessário estudar outras distribuições circulares para que possamos reconhecer a distribuição adequada para as observações. Portanto, para trabalhos futuros, a pretensão é elaborar uma família de distribuições circulares como tentativa de abranger possíveis conjunto de dados obtendo resultados eficientes e precisos.

REFERÊNCIAS

- ABEYASEKERA, S.; COLLETT, D. On the estimation of the parameters of the von mises distribution. **Communications in Statistics - Theory and Methods**, Taylor Francis, v. 11, n. 18, p. 2083–2090, 1982. Disponível em: <<https://doi.org/10.1080/03610928208828371>>.
- AGOSTINELLI, C.; LUND, U. **R package circular: Circular Statistics (version 0.4-93)**. CA: Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, Italy. UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA, 2017. Disponível em: <<https://r-forge.r-project.org/projects/circular/>>.
- ARTES, R.; PAULA, G. Extensões da teoria das equações de estimação generalizadas a dados circulares e modelos de dispersão /. 1997.
- BATSCHELET. **Circular Statistics in Biology**. [S.l.]: London: Academic Press, 1981.
- BERGIN, T. A comparison of goodness-of-fit tests for analysis of nest orientation in western kingbirds (*tyrannus verticalis*). **Condor** **93:164–171**, 1991.
- COLEMAN, D.; HASKEY, J. Marital distance and its geographical orientation in england and wales, 1979. **Transactions (Institute of British Geographers : 1965)**, v. 11, p. 337–55, 02 1986.
- DHEERU, D.; TANISKIDOU, E. K. **UCI Machine Learning Repository**. 2017. Disponível em: <<http://archive.ics.uci.edu/ml>>.
- DOWNS e GOULD. **Some relationships between the normal and von Mises distribution**. [S.l.]: Biometrika, 1967.
- FISHER. **Statistical Analysis of Circular Data**. [S.l.]: Cambridge University Press., 1993.
- GUMBEL, E. J. **Applications of the circular normal distribution**. [S.l.]: J. Amer, 1954.
- GUMBEL, E. J, GREENWOOD, J. A E DURAND, D. . **The circular normal distribution: theory and tables**. [S.l.]: J. Amer, 1953.
- JAMMALAMADAKA; SENGUPTA. **Topics in Circular Statistics**. 3. ed. : World Scientific Publishing Co. Pte. Ltd., 2001. v. 5. (, v. 5). pag, 79-81. ISBN .
- LEAL. Análise de resíduos em modelos de regressão von mises. **Pós-Graduação em Matemática - CCT - UFCG**, 2006.
- LUKACS. E. **Characteristic Functions**. [S.l.]: Griffin, London, 1970.
- MARDIA. **Statistics of Directional Data**. [S.l.]: London: Academic Press., 1972.
- MURALIDHARAN, K.; PARIKH, d. r. Some bayesian inferences for von mises distribution. **American Journal of Mathematical and Management Sciences**, v. 27, 01 2007.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2013. Disponível em: <<http://www.R-project.org/>>.
- RAO. J. S. **Some Contributions to the Analysis of Circular data**. [S.l.]: Ph. D. thesis. Indian Stat, 1969.

ROBERT, T. M. Gray e M. **A Treatise on Bessel Function**. [S.l.: s.n.], 1922.

SOUZA, F. D. Influência local e análise de resíduos em modelos de regressão von mises. **Instituto de Matemática e Estatística - USP**, 1999.

SOUZA, F. D.; PAULA, G. Deviance residuals for an angular response. **Australian and New Zealand Journal of Statistics**, p. 385–390, 09 2002.

SPURR, B. D.; KOUTBEIY, M. A. A comparison of various methods for estimating the parameters in mixtures of von mises distribution. **Communications in Statistics-simulation and Computation - COMMUN STATIST-SIMULAT COMPUT**, v. 20, p. 725–741, 01 1991.

USHAKOV, N. G. **Selected Topics in Characteristic Functions**. [S.l.]: Utrecht, The Netherlands, 1999.

WATSON, G. S. **Some problems in the statistics of directions**. [S.l.]: Bull of the *36th* Session, 1967.

APÊNDICE A – CÓDIGOS PARA OS GRÁFICOS E MEDIDAS DESCRITIVAS

PROVA:

$$\begin{aligned}
 \frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \bar{x}_o) &= \frac{1}{n} \sum_{i=1}^n (\cos \theta_j \cos \bar{x}_o + \text{sen} \theta_j \text{sen} \bar{x}_o) \\
 &= \bar{C} \cos(\bar{x}_o) + \bar{S} \text{sen}(\bar{x}_o) \\
 &= \bar{R} \cos^2(\bar{x}_o) + \bar{R} \text{sen}^2(\bar{x}_o) \\
 &= \bar{R}.
 \end{aligned}$$

PROVA:

$$\begin{aligned}
 \sum_{i=1}^n \cos(\theta_i - \mu) &= \sum_{i=1}^n (\cos(\theta_i) \cos(\mu) + \text{sen}(\theta_i) \text{sen}(\mu)) \\
 &= \cos(\mu) \sum_{i=1}^n \cos(\theta_i) + \text{sen}(\mu) \sum_{i=1}^n \text{sen}(\theta_i) \\
 &= \cos(\mu) \left(\frac{n}{n} \sum_{i=1}^n \cos(\theta_i) \right) + \text{sen}(\mu) \left(\frac{n}{n} \sum_{i=1}^n \text{sen}(\theta_i) \right) \\
 &= \cos(\mu) n \bar{C} + \text{sen}(\mu) n \bar{S} \\
 &= \cos(\mu) n \bar{R} \cos(\bar{x}_o) + \text{sen}(\mu) n \bar{R} \text{sen}(\bar{x}_o) \\
 &= n \bar{R} (\cos(\mu) \cos(\bar{x}_o) + \text{sen}(\mu) \text{sen}(\bar{x}_o)) \\
 &= n \bar{R} \cos(\bar{x}_o - \mu)
 \end{aligned}$$

```

install.packages("circular")
install.packages("NISTunits", dependencies = TRUE)
install.packages("xtable")
library(xtable)
library(circular)
library(NISTunits)

## Gráfico da densidade
x = seq(-pi, pi, by = 0.01)
plot(x = x, y = dvonmises(x,circular(0),0),type="l",ylim=c(0,1),
ylab = "Função densidade", xaxt="none")
axis(side=2)
axis(side=1, at=c(-pi, -pi/2, -pi/4, 0 ,pi/4, pi/2, pi))

```

```

labels=expression(-pi, -pi/2, -pi/4, 0, pi/4, pi/2, pi))
lines(x = x, y = dvonmises(x,circular(0),1), col = "red", add=T)
lines(x = x, y = dvonmises(x,circular(0),2), col = "green", add=T)
lines(x = x, y = dvonmises(x,circular(0),4),col = "blue", add=T)
legend(2, 0.9, legend=c(expression(paste(kappa, " = ", 0)),
                        expression(paste(kappa, " = ", 1)),
                        expression(paste(kappa, " = ", 2)),
                        expression(paste(kappa, " = ", 4))),
      col=c("black", "red", "green", "blue"), lty=c("solid", "solid",
      "solid","solid"), cex=0.8, bty="n")

```

```
## Gráfico da acumulada
```

```

x = seq(-pi, pi, by = 0.01)
plot(x = x, y = pvonmises(x,circular(0),0),type="l",ylim=c(0,1),
ylab = "Função acumulada", xaxt="none")
axis(side=2)
axis(side=1, at=c(-pi, -pi/2, -pi/4, 0 ,pi/4, pi/2, pi))
labels=expression(-pi, -pi/2, -pi/4, 0, pi/4, pi/2, pi))
lines(x = x, y = pvonmises(x,circular(0),1), col = "red", add=T)
lines(x = x, y = pvonmises(x,circular(0),2), col = "green", add=T)
lines(x = x, y = pvonmises(x,circular(0),4),col = "blue", add=T)
legend(2, 0.5, legend=c(expression(paste(kappa, " = ", 0)),
                        expression(paste(kappa, " = ", 1)),
                        expression(paste(kappa, " = ", 2)),
                        expression(paste(kappa, " = ", 4))),
      col=c("black", "red", "green", "blue"),
      lty=c("solid", "solid", "solid","solid"),
      cex=0.8, bty="n")

```



```
## Média direcional para dados agrupados
```

```
mediacirc=function(x, f = 1){
```

```
  mu=0
```

```
  s=sum(f*sin(x))
```

```
  c=sum(f*cos(x))
```

```
  arc=atan(s/c)
```

```
  if(s>=0 && c>0){
```

```
    mu=arc
```

```
  }else {
```

```
    if(c<0){
```

```
      mu=arc+pi
```

```
    }else{
```

```
      if(s<0 && c>0){
```

```
        mu=arc+2*pi
```

```
      }
```

```
    }
```

```
  }
```

```
  return(mu)
```

```
}
```

```
dados = c(10,30,50,70,90,110,130,150,170,190,210,230,250,270,290,310,330,350)
```

```
f = c(40,22,20,9,6,3,3,1,6,3,11,22,24,58,136,138,143,69)
```

```
media.circ(dados,f)
```

```
[1] 314.2704
```

```
## Simulação
n = 1000
mu = kappa = NULL
m = 0.2
k = 4
for (i in 1:1000) {
  set.seed(1003*i + 1997*n)
  x = rvonmises(n, circular(m), k)
  mu = c(mu, mean.circular(x))
  kappa = c(kappa, A1inv(mean(cos(x - mean.circular(x)))))
}
c(mean(mu), mean(mu - m), mean((mu - m)^2))
c(mean(kappa), mean(kappa - k), mean((kappa - k)^2))
```

APÊNDICE B – CÓDIGOS UTILIZADOS NA APLICAÇÃO

```

install.packages("circular")
install.packages("ggplot2")
library(ggplot2)
library(circular)
data(fisherB13c)
str(fisherB13c)

## Análise descritiva linear
n = medial = medianal = varl = dpl = 0
for (i in 1:14) {
  n[i] = length(fisherB13[[i]])
  medial[i] = mean(fisherB13[[i]])
  varl[i] = var(fisherB13[[i]])
  dpl[i] = sd(fisherB13[[i]])
  medianal[i] = median(fisherB13[[i]])
}
xtable(cbind(n,medial,medianal,varl,dpl))

## Análise descritiva angular
n = media = amp = var = dp = r = trigm = 0
for (i in 1:14) {
  n[i] = length(fisherB13c[[i]])
  s=sum(sin(fisherB13c[[i]]))/n
  c=sum(cos(fisherB13c[[i]]))/n
  r[i] = sqrt(c^2 + s^2)
  media[i] = mean.circular(fisherB13c[[i]])
  amp[i] = range.circular(fisherB13c[[i]])
  var[i] = var.circular(fisherB13c[[i]])
  dp[i] = sd.circular(fisherB13c[[i]])
  trigm[i] = trigonometric.moment(fisherB13c[[i]], p = 2)
}

```

```

## Tabela - linear
xtable(round(cbind(fisherB13$set1[1:6],fisherB13$set2[1:6],fisherB13$set3[1:6],
fisherB13$set4[1:6],fisherB13$set5[1:6],fisherB13$set6[1:6],
fisherB13$set7[1:6]),0))

xtable(round(cbind(fisherB13$set9[1:6],fisherB13$set10[1:6],fisherB13$set11[1:6],
fisherB13$set12[1:6],fisherB13$set13[1:6],fisherB13$set14[1:6],0))

## Tabela - angular
tr <- as.numeric(trigm)
f = round(rbind(n,r,media,amp,var,dp,tr),4)
colnames(f) <- c("Local 1", "Local 2", "Local 3", "Local 4", "Local 5", "Local 6",
"Local 7", "Local 8", "Local 9", "Local 10", "Local 11", "Local 12",
"Local 13", "Local 14")
xtable(f[,1:7])
xtable(f[,8:14])

## Gráficos
par(mfrow=c(1,2))
plot(fisherB13c$set1,stack=TRUE, units="degrees", bins = 100, shrink=1.5)
rose.diag(fisherB13c$set1, bins=20, units="degrees",col="darkgrey",add=TRUE)
pp.plot(fisherB13c$set1, xlab = "Distribuição Normal Circular",
ylab = "Distribuição Empírica")

plot(fisherB13c$set9,stack=TRUE, units="degrees", bins = 100, shrink=1.5)
rose.diag(fisherB13c$set9, bins=16, units="degrees",col="darkgrey",add=TRUE)
pp.plot(fisherB13c$set9, xlab = "Distribuição Normal Circular",
ylab = "Distribuição Empírica")

```

```

## Histograma
qplot(fisherB13$set1, geom="histogram", xlab = "Localidade 1",
ylab = "Frequência",xlim = c(0,300), binwidth = 0.5) +
geom_histogram(bins = 12,color="darkblue", fill="lightblue")

qplot(fisherB13$set9, geom="histogram", xlab = "Localidade 9",
ylab = "Frequência", xlim = c(120, 220), ylim = c(0,30),
binwidth = 0.1) + geom_histogram(bins = 11, color="darkblue", fill="lightblue")

## Teste Watson
watson.test(fisherB13c$set1, alpha = 0.05, dist = "vonmises")
watson.test(fisherB13c$set9, alpha = 0.05, dist = "vonmises")

##### import column_2C
length(X3); mean(X3); median(X3); var(X3); sd(X3)

X3 = circular(X3, units = "degrees")
media = mean.circular(X3); media
variancia = var.circular(X3); variancia
trigo = trigonometric.moment(X3, p = 2); trigo
desvio = sd.circular(X3); desvio
amplitude = range.circular(X3); amplitude

## Dispersão , diagrama de rosas e PP-plot
par(mfrow = c(1,2))
plot(X3,stack=TRUE, units="degrees", bins = 100, shrink=1.5)
rose.diag(X3, bins=16, units="degrees",col="darkgrey",add=TRUE)
pp.plot(X3, xlab = "Distribuição Normal Circular",
ylab = "Distribuição Empírica")

## Teste Watson
watson.test(X3, alpha = 0.05, dist = "vonmises")

```