



Universidade Federal
de São João del-Rei



UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL REI
PROFMAT - MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL

ELTON WEIGLAS DE PAULA

ESTATÍSTICA CIRCULAR: UM ESTUDO
DIRECIONAL, VIA COORDENADAS
GEOGRÁFICAS, DOS ALUNOS
APROVADOS NA ESCOLA
PREPARATÓRIA DE CADETES DO AR

SÃO JOÃO DEL REI
2023

ELTON WEIGLAS DE PAULA

**ESTATÍSTICA CIRCULAR: UM ESTUDO DIRECIONAL, VIA
COORDENADAS GEOGRÁFICAS, DOS ALUNOS APROVADOS NA
ESCOLA PREPARATÓRIA DE CADETES DO AR**

Dissertação apresentada à Universidade Federal de São João del Rei, Campus Santo Antônio, como parte das exigências do Programa de Pós-Graduação Mestrado Profissional em Matemática em Rede Nacional, para obter o título de Mestre.

Orientadora

Profa. Dra. Carla Regina Guimarães Brighenti -
UFSJ

SÃO JOÃO DEL REI
2023

Ficha catalográfica elaborada pela Divisão de Biblioteca (DIBIB)
e Núcleo de Tecnologia da Informação (NTINF) da UFSJ,
com os dados fornecidos pelo(a) autor(a)

P324e Paula, Elton Weiglas de.
Estatística Circular: Um estudo direcional, via coordenadas geográficas, dos alunos aprovados na Escola Preparatória de Cadetes do Ar / Elton Weiglas de Paula ; orientadora Carla Regina Guimarães Brighenti. -- São João del-Rei, 2023.
47 p.

Dissertação (Mestrado - Programa de Mestrado Profissional em Matemática em Rede Nacional - PROFMAT) -- Universidade Federal de São João del-Rei, 2023.

1. estatística descritiva circular. 2. média direcional. 3. desvio padrão circular. 4. boxplot circular. 5. distribuição von Mises. I. Brighenti, Carla Regina Guimarães, orient. II. Título.

ELTON WEIGLAS DE PAULA

**ESTATÍSTICA CIRCULAR: UM ESTUDO DIRECIONAL, VIA
COORDENADAS GEOGRÁFICAS, DOS ALUNOS APROVADOS NA
ESCOLA PREPARATÓRIA DE CADETES DO AR**

Dissertação apresentada à Universidade Federal de São João del Rei, Campus Santo Antônio, como parte das exigências do Programa de Pós-Graduação Mestrado Profissional em Matemática em Rede Nacional, para obter o título de Mestre.

APROVADA: 26 de abril de 2023.

Profa. Dra. Carla Regina Guimarães Brighenti - UFSJ
(Orientadora)

Profa. Dra. Viviane Pardini Valério - UFSJ
(Membro Interno do PROFMAT)

Prof. Dr. Clodoaldo Teodosio Santana da Silva - UFVJM
(Membro Externo)

SÃO JOÃO DEL REI
2023

Dedico este trabalho à minha querida esposa
Paloma e aos meus filhos Gabriel e Maria
Fernanda.

Agradecimentos

Primeiramente, agradeço a Deus, por me conceder o dom da vida e por me proporcionar momentos incríveis.

À minha família que sempre esteve ao meu lado na realização deste sonho, meu pai Elson, minha mãe Rosa, meus irmãos Maicon e Felipe por me acompanharem nas decisões e principalmente pela nossa amizade e amor.

À minha esposa Paloma, pela sua imensa compreensão das minhas ausências para me dedicar aos estudos, mas mesmo assim, me deu todo o suporte que precisava nos momentos de fraqueza, sem ela eu não conseguiria.

Aos meus amados filhos, Gabriel e Maria Fernanda, por serem não só minha força diária mas também minha motivação.

À minha orientadora Profa. Carla Brighenti pela paciência, acolhimento, comprometimento e dedicação em me orientar. Meu muito obrigado por tudo que me ensinou.

A Profa. Viviane e ao Prof. Clodoaldo por aceitarem o convite em compor a banca e também em contribuir com suas sugestões em busca de um trabalho melhor.

Agradeço a todos os professores do PROFMAT/CSA-UFSJ pelos valiosos ensinamentos passados.

Aos meus colegas de turma, Edilamar, Natália, Rodney e Rogério, que me acompanharam e somaram para que as nossas sextas-feiras fossem mais prazerosas.

Aos meus amigos de trabalho, por abraçarem minha decisão, por me apoiarem e entenderem minhas ausências.

Enfim, agradeço a todos que, direta ou indiretamente, me ajudaram na conclusão deste mestrado.

Resumo

A estatística circular tem aplicações em diversas áreas cujos dados possuam direções, ângulos ou tenham comportamento periódico, tais como: direção de migração de animais, dos ventos ou de materiais geológicos, horários de pico de trânsito ou de nascimento de inseto, além de coordenadas geográficas entre outros. No entanto, em nosso cotidiano podemos avaliar instintivamente, de forma equivocada, dados com características angulares de forma clássica, utilizando a estatística descritiva linear. Assim, o objetivo deste trabalho foi estudar a teoria da estatística circular e comparar os resultados entre a estatística descritiva linear, ou seja, utilizada para dados na reta real, contemplada no ensino regular e a estatística circular, evidenciando a importância do estudo desta para dados angulares e apresentar gráficos circulares, com o auxílio do software R. Para aplicação da teoria será utilizado um conjunto de dados reais das coordenadas geográficas das cidades de origem de alunos da Escola Preparatória de Cadetes do Ar (EPCAR) comparando as medidas estatísticas de três diferentes anos de ingresso.

Palavras-chave: estatística descritiva circular. média direcional. desvio padrão circular. boxplot circular. distribuição von Mises.

Abstract

Circular statistics has applications in several areas whose data have directions, angles or have periodic behavior, such as: direction of migration of animals, winds or geological materials, times of peak traffic or insect birth, in addition to geographic coordinates between others. However, in our daily lives, we can instinctively evaluate, in a wrong way, data with angular characteristics in a classical way, using linear descriptive statistics. Thus, the objective of this work was to study the theory of circular statistics and compare the results between linear descriptive statistics, that is, used for data on the real line, contemplated in regular education, and circular statistics, evidencing the importance of studying the latter for data angular and present circular graphics, with the aid of the R software. For application of the theory, a set of real data of the geographic coordinates of the cities of origin of students of the Preparatory School for Air Cadets (EPCAR) will be used, comparing the statistical measures of three different entry years.

Keywords: circular descriptive statistics. directional average. circular standard deviation. circular boxplot. von Mises distribution.

Lista de Símbolos

Símbolos e notações utilizadas neste trabalho:

R	Comprimento do vetor resultante.	Φ	Mediana angular.
\bar{R}	Comprimento médio do vetor resultante.	\bar{x}	Média linear.
v	Desvio padrão circular.	Me	Mediana linear.
V	Variância circular.	s	Desvio padrão linear.
$\bar{\theta}$	Média direcional.	s^2	Variância linear.
κ	Parâmetro de concentração.		

Lista de Abreviaturas

Abreviaturas utilizadas neste trabalho:

AFA	Academia da Força Aérea
BNCC	Base Nacional Curricular Comum
CFOAV	Curso de Formação de Oficiais Aviadores
CPCAR	Curso Preparatório de Cadetes do Ar
CSA	Campus Santo Antônio
DLA	Diferença de latitude
DLO	Diferença de longitude
EPCAR	Escola Preparatória de Cadetes do Ar
IBGE	Instituto Brasileiro de Geografia e Estatística
OMS	Organização Mundial de Saúde
PROFMAT	Mestrado Profissional em Matemática em Rede Nacional
UFSJ	Universidade Federal de São João del Rei

Lista de Figuras

2.1	Representação dos pontos P e Q.	14
2.2	Representação de um ponto P.	15
2.3	Gráficos das funções $y = \text{tg}(x)$ e $x = \text{arctg}(y)$ respectivamente.	17
2.4	Direção do voo, em graus, das 11 rainhas.	20
2.5	Mediana direcional do primeiro voo, em graus, das 11 rainhas, representada pelo gráfico dos pontos angulares e pelo Boxplot circular.	22
2.6	Gráfico da função densidade de probabilidade von Mises das 11 rainhas. . .	27
3.1	Coordenadas geográficas de Barbacena e Belo Horizonte em um sistema de eixos cartesianos	30
3.2	Distância em linha reta entre Barbacena e Belo Horizonte	31
3.3	Gráfico de pontos das turmas ingressantes nos anos de 2020, 2021 e 2022 respectivamente.	36
3.4	Diagrama de pontos das turmas ingressantes nos anos de 2020, 2021 e 2022 respectivamente.	37
3.5	Diagrama de rosas, boxplot circular e linear da turma ingressante no ano de 2020.	38
3.6	Diagrama de rosas, boxplot circular e linear da turma ingressante no ano de 2021.	38
3.7	Diagrama de rosas, boxplot circular e linear da turma ingressante no ano de 2022.	38
4.1	Gráfico da função densidade de probabilidade von Mises das 3 turmas. . .	41

Lista de Tabelas

2.1	Direção do voo nupcial, em graus, das 11 rainhas.	19
2.2	Comparação dos resultados do exemplo do voo das 11 abelhas rainhas obtidos através da estatística linear e angular.	23
3.1	Estados de origem dos alunos do 3 ^o esquadrão da EPCAR, ingressantes no ano de 2020 e suas respectivas coordenadas geográficas considerando as distâncias das capitais em relação à Barbacena - MG.	32
3.2	Parâmetros dos dados circulares e lineares dos alunos do 3 ^o esquadrão da EPCAR, ingressantes no ano de 2020.	32
3.3	Estados de origem dos alunos do 2 ^o esquadrão da EPCAR, ingressantes no ano de 2021 e suas respectivas coordenadas geográficas considerando as distâncias das capitais em relação à Barbacena - MG.	33
3.4	Parâmetros dos dados circulares e lineares dos alunos do 2 ^o esquadrão da EPCAR, ingressantes no ano de 2021.	34
3.5	Estados de origem dos alunos do 1 ^o esquadrão da EPCAR, ingressantes no ano de 2022 e suas respectivas coordenadas geográficas considerando as distâncias das capitais em relação à Barbacena - MG.	35
3.6	Parâmetros dos dados circulares e lineares dos alunos do 1 ^o esquadrão da EPCAR, ingressantes no ano de 2022.	35

Sumário

1	INTRODUÇÃO	12
2	ESTATÍSTICA DESCRITIVA CIRCULAR	14
2.1	Média Direcional	15
2.2	Mediana Circular	21
2.3	Medidas de Dispersão	22
2.4	Distribuição de Probabilidade Circular	23
2.4.1	Distribuição Uniforme Circular	24
2.4.2	Distribuição de von Mises	24
2.5	Testes de hipóteses	25
3	USO DA ESTATÍSTICA CIRCULAR PARA DADOS GEOGRÁFICOS	28
4	DISCUSSÕES	39
5	CONSIDERAÇÕES FINAIS	42
	REFERÊNCIAS	44
	APÊNDICES	
A	Comandos utilizados no software R	45

1 INTRODUÇÃO

Observa-se em diversas áreas de conhecimento um enorme interesse em estudar fenômenos que fornecem dados direcionais ou medidas angulares. De fato, quando estamos interessados em estudar fenômenos cujos dados possuem direções, tais como: direção dos ventos, transporte de materiais geológicos, direção de migrações de animais, entre outros a ferramenta ideal é denominada estatística circular.

Os dados direcionais podem ser representados geometricamente por pontos em uma circunferência centrada na origem do sistema do plano cartesiano e pode-se também utilizar a representação em coordenadas polares, observe que quando se fixa um determinado ponto P no plano cartesiano cujas coordenadas são (x,y) , esse ponto P possui uma distância r em relação à origem, ao assumir um vetor que esteja no eixo das abscissas e, ao rotacionar este vetor no sentido anti-horário até que o mesmo passe por P, o deslocamento formado entre o vetor e o eixo x definirá um ângulo θ , desta maneira P poderá ser retratado como $(r \cos \theta, r \sin \theta)$, assim é possível analisar cada dado disponível de forma angular.

Constata-se ainda que, ao analisar dados direcionais e o interesse esta apenas em estudar a direção e não a magnitude do vetor, a análise é facilitada pela representação através de vetores de comprimento unitário.

Em especial, um conjunto de dados que apresenta coordenadas geográficas pode ser transformado e analisado como medidas angulares. No entanto, há uma predisposição em querer manipular esses dados circulares com as ferramentas da estatística descritiva linear o que resultaria em soluções não confiáveis, alguns autores apresentam o uso da estatística circular para exemplificar tal conjunto de dados [1] [2] [3].

O ensino da estatística descritiva linear está previsto na Base Nacional Curricular Comum - (BNCC), no item 5.2.1.1, que trata sobre as considerações sobre a organização curricular para o ensino médio, segundo habilidade (EM13MAT316), é estabelecido como meta para o ensino de estatística que os alunos possuam um domínio em "Resolver e elaborar problemas,

em diferentes contextos, que envolvem cálculo e interpretação das medidas de tendência central (média, moda, mediana) e das medidas de dispersão (amplitude, variância e desvio padrão)”[4], bem como análise e interpretações de gráficos. Particularmente, as médias que são discutidas, geralmente, nas salas de aulas são as médias aritmética, geométrica, harmônica. No entanto, não há nenhuma menção sobre as médias de dados angulares, apesar de sua extensa aplicabilidade.

Diante do exposto, este trabalho visa apresentar o uso da estatística circular como ferramenta para análise de dados angulares e seu uso aplicado no conjunto de dados dos alunos da Escola Preparatória de Cadetes do Ar (EPCAR).

Inicialmente, será apresentado uma revisão sobre os conceitos básicos no âmbito da estatística circular, tais como medidas descritivas, distribuição de probabilidade e testes de hipóteses, juntamente com a exibição de medidas de posição e dispersão advindas da estatística linear para fins de comparação entre as duas áreas de estudo.

Posteriormente, esses conceitos serão utilizados em um estudo direcional sobre a procedência dos estudantes que ingressaram na EPCAR entre os anos de 2020 à 2022.

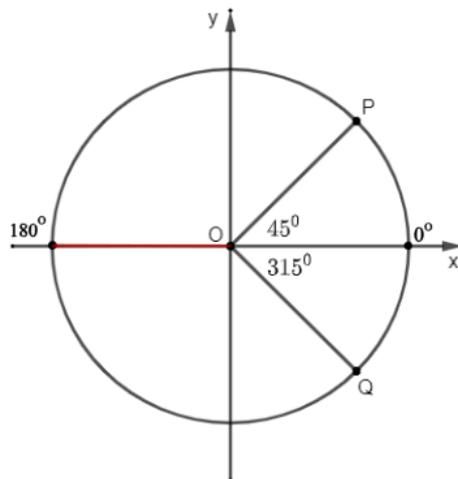
2 ESTATÍSTICA DESCRITIVA CIRCULAR

Para ilustrar o problema abordado neste trabalho, pode-se avaliar uma situação simples dada por uma amostra com apenas dois dados angulares, como por exemplo $P = 45^\circ$ e $Q = 315^\circ$. Ao calcular a média aritmética usual desses dados obtém-se o seguinte resultado:

$$\bar{x} = \frac{45^\circ + 315^\circ}{2} = 180^\circ$$

Contudo, ao plotar tais dados na circunferência, nota-se que a média aritmética 180° não condiz com o esperado da Figura(2.1), sendo que o valor mais intuitivo seria a média igual a 0° . Assim percebe-se a necessidade de utilizar outro recurso estatístico para obter medidas mais condizentes.

Figura 2.1: Representação dos pontos P e Q.

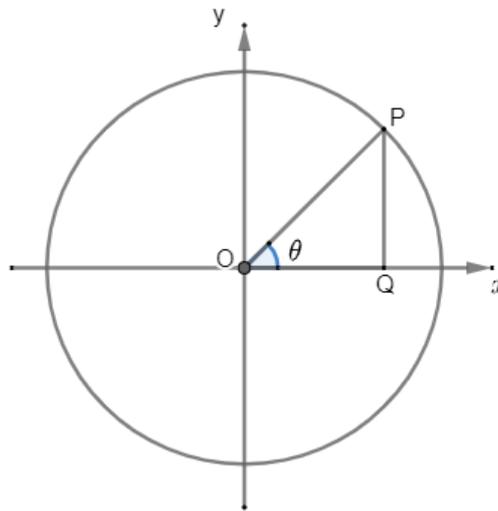


Fonte: Os Autores.

Seja (x,y) coordenadas de um ponto P no plano cartesiano, observe que este mesmo ponto $P = (x,y)$ pode ser representado sobre uma circunferência de centro $O = (0,0)$ e

raio r , pois a distância do ponto P ao centro O , denotada por $d(O,P) = r$, satisfaz as condições de P pertencer a circunferência (Figura 2.2).

Figura 2.2: Representação de um ponto P .



Fonte: Os Autores.

Note que $\overline{OP} = r$ e que o menor ângulo formado por \overline{OP} com o eixo x é θ , então aplicando as relações métricas de um triângulo retângulo obtemos

$$\cos(\theta) = \frac{x}{r} \Leftrightarrow x = r \cos \theta \quad (2.1)$$

e

$$\text{sen}(\theta) = \frac{y}{r} \Leftrightarrow y = r \text{sen} \theta \quad (2.2)$$

assim o ponto $P = (x,y)$ pode ser representado por $P = (r \cos \theta, r \text{sen} \theta)$. Para $r = 1$, tem-se que o ponto P pertence a um círculo unitário caracterizado por $P = (\cos \theta, \text{sen} \theta)$.

2.1 Média Direcional

Considere, uma amostra aleatória de tamanho n dada pelos pontos $P_1, P_2, P_3, \dots, P_n$ em um círculo unitário, e seus respectivos ângulos $\theta_1, \theta_2, \theta_3, \dots, \theta_n$ em relação ao eixo x . Cada ponto P_i pode ser representado por um vetor $P_i = (\cos \theta_i, \text{sen} \theta_i)$ em que o comprimento

R do vetor resultante dos n vetores unitários, será dado por

$$R = \sqrt{\left(\sum_{i=1}^n \cos \theta_i\right)^2 + \left(\sum_{i=1}^n \sen \theta_i\right)^2} \quad (2.3)$$

Assumindo $C = \sum_{i=1}^n \cos \theta_i$ e $S = \sum_{i=1}^n \sen \theta_i$, o comprimento do vetor resultante será descrito pela seguinte igualdade:

$$R = \sqrt{C^2 + S^2}. \quad (2.4)$$

Em seguida, será apresentado o conceito da **Média Direcional** ($\bar{\theta}$), sendo que esta média tem a mesma direção que o centro de massa do círculo unitário [5].

O centro de massa será dado por $M = (\bar{C}, \bar{S})$ tal que

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n \cos \theta_i \quad (2.5)$$

$$\bar{S} = \frac{1}{n} \sum_{i=1}^n \sen \theta_i \quad (2.6)$$

Define-se também o **comprimento médio resultante** (\bar{R}) como sendo o tamanho médio do vetor resultante e este é obtido pela seguinte equação.

$$\bar{R}^2 = \left(\frac{1}{n} \sum_{i=1}^n \cos \theta_i\right)^2 + \left(\frac{1}{n} \sum_{i=1}^n \sen \theta_i\right)^2 \quad (2.7)$$

$$\bar{R}^2 = \bar{C}^2 + \bar{S}^2 \Rightarrow \bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2} \quad (2.8)$$

em que $0 \leq \bar{R} \leq 1$, pois $\bar{R} = \frac{R}{n}$.

O comprimento médio resultante é dito como uma medida de dispersão, pois ele fornece um grau de concentração dos ângulos observados e, quanto mais próximo de 1 o valor de \bar{R} está, menos dispersos estão os pontos. Por outro lado, quanto mais próximo de 0 for este valor \bar{R} , menos concentrados estão os dados [1].

Deste modo, temos que a **Média Direcional** é definida como o ângulo $(\bar{\theta})$ e deve satisfazer as seguintes equações:

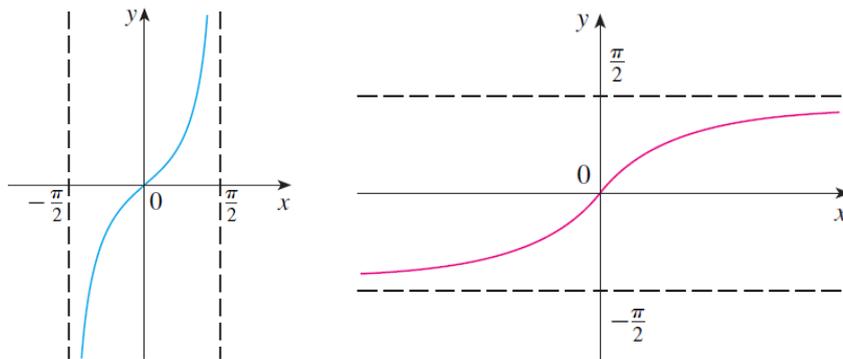
$$\begin{cases} \cos \bar{\theta} = \frac{(\bar{C})}{(\bar{R})} \\ \text{sen } \bar{\theta} = \frac{(\bar{S})}{(\bar{R})} \end{cases} \quad (2.9)$$

como

$$\text{tg}(\bar{\theta}) = \frac{\text{sen}(\bar{\theta})}{\cos(\bar{\theta})} = \frac{\frac{(\bar{S})}{(\bar{R})}}{\frac{(\bar{C})}{(\bar{R})}} \Leftrightarrow \text{tg}(\bar{\theta}) = \frac{(\bar{S})}{(\bar{C})}. \quad (2.10)$$

e, levando-se em consideração que a função $f(x) = \text{tg}(x)$, restrita ao domínio cujo intervalo é $f : \left] \frac{-\pi}{2}, \frac{\pi}{2} \right[\rightarrow \mathbb{R}$, tem-se como consequência que $f(x)$ é bijetiva. Assim, a função f admite inversa, denominada por arcotangente e denotada por $f^{-1} : \mathbb{R} \rightarrow \left] \frac{-\pi}{2}, \frac{\pi}{2} \right[$.

Figura 2.3: Graficos das funções $y = \text{tg}(x)$ e $x = \text{arctg}(y)$ respectivamente.



Fonte: Os autores.

Ou seja, como a função arctg é definida apenas no primeiro e quarto quadrante, é necessário uma análise em relação aos sinais do \bar{S} e \bar{C} para que a média angular seja dada de forma fiel em relação aos demais quadrantes. Ao examinar as possibilidades dos sinais, obtem-se o seguinte:

1. Se $\bar{S} > 0$ e $\bar{C} > 0 \Rightarrow 0 \leq \text{arctg} \frac{(\bar{S})}{(\bar{C})} < \frac{\pi}{2}$;

$$2. \text{ Se } \bar{S} > 0 \text{ e } \bar{C} < 0 \Rightarrow \frac{\pi}{2} < \operatorname{arctg} \frac{(\bar{S})}{(\bar{C})} \leq \pi;$$

$$3. \text{ Se } \bar{S} < 0 \text{ e } \bar{C} < 0 \Rightarrow \pi < \operatorname{arctg} \frac{(\bar{S})}{(\bar{C})} < \frac{3\pi}{2};$$

$$4. \text{ Se } \bar{S} < 0 \text{ e } \bar{C} > 0 \Rightarrow \frac{3\pi}{2} < \operatorname{arctg} \frac{(\bar{S})}{(\bar{C})} \leq 2\pi.$$

portanto, o valor da média direcional será obtido pelas seguintes expressões:

$$\bar{\theta} = \begin{cases} \operatorname{arctg} \frac{(\bar{S})}{(\bar{C})}, & \text{se } \bar{S} > 0 \text{ e } \bar{C} > 0 \\ \operatorname{arctg} \frac{(\bar{S})}{(\bar{C})} + \pi, & \text{se } \bar{C} < 0 \\ \operatorname{arctg} \frac{(\bar{S})}{(\bar{C})} + 2\pi, & \text{se } \bar{S} < 0 \text{ e } \bar{C} > 0. \end{cases} \quad (2.11)$$

Para exemplificar esta parte teórica e retomando ao exemplo fictício citado anteriormente com dois dados angulares segue que sua média circular pode ser encontrada através dos seguintes passos:

Primeiramente, calcula-se o centro de massa M dos pontos $P = 45^\circ$ e $Q = 315^\circ$ segundo as equações (2.5) e (2.6).

$$\begin{aligned} \bar{C} &= \frac{1}{2} \sum_{i=1}^2 \cos \theta_i \\ &= \frac{1}{2} (\cos (45^\circ) + \cos (315^\circ)) \\ &= \frac{1}{2} \left(\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} \right) \\ &= \frac{\sqrt{2}}{2} \end{aligned}$$

$$\begin{aligned} \bar{S} &= \frac{1}{2} \sum_{i=1}^2 \operatorname{sen} \theta_i \\ &= \frac{1}{2} (\operatorname{sen} (45^\circ) + \operatorname{sen} (315^\circ)) \\ &= \frac{1}{2} \left(\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2} \right) \\ &= 0 \end{aligned}$$

deste modo, o comprimento médio do vetor resultante conforme (2.8) é

$$\begin{aligned}\bar{R} &= \sqrt{\bar{C}^2 + \bar{S}^2} \\ &= \sqrt{\left(\frac{\sqrt{2}}{2}\right)^2 + (0)^2} \\ &= \frac{\sqrt{2}}{2}\end{aligned}$$

assim, a média direcional (2.9) será

$$\begin{cases} \cos \bar{\theta} = \frac{(\bar{C})}{(\bar{R})} = \frac{\frac{\sqrt{2}}{2}}{\frac{\sqrt{2}}{2}} = 1 \\ \text{sen } \bar{\theta} = \frac{(\bar{S})}{(\bar{R})} = \frac{0}{\frac{\sqrt{2}}{2}} = 0 \end{cases}$$

portanto,

$$\begin{aligned}\bar{\theta} &= \text{arctg} \frac{(\bar{S})}{(\bar{C})} \\ &= \frac{0}{1} = 0^\circ.\end{aligned}$$

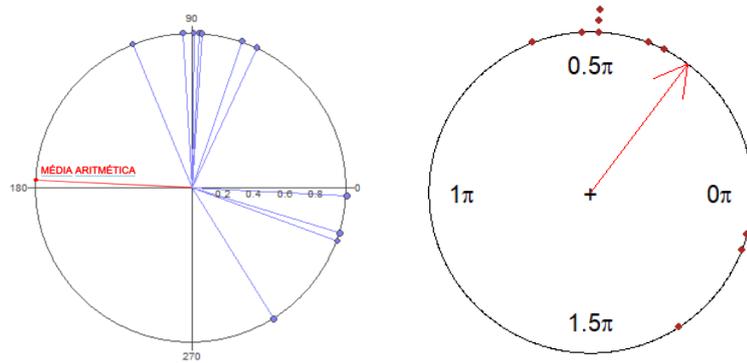
Logo, a média circular $\bar{\theta}$ dos ângulos 45° e 315° é igual a 0° . Para ampliar o entendimento do assunto, apresenta-se a seguir um exemplo adaptado de [6]:

Exemplo: Em colmeias de abelhas foi realizado um estudo sobre a direção do voo nupcial de 11 rainhas, em relação a entrada da colmeia obtendo-se os seguintes resultados conforme Tabela (2.1). Em seguida, estão plotados a direção do voo dessas rainhas bem como a direção da média aritmética linear (gráfico a esquerda) e a direção média circular (gráfico a direita).

Tabela 2.1: Direção do voo nupcial, em graus, das 11 rainhas.

Rainha	1	2	3	4	5	6	7	8	9	10	11
Ângulos	87	343	71	112	86	340	357	65	93	302	89

Fonte: Adaptado de Ayres (2007).

Figura 2.4: Direção do voo, em graus, das 11 rainhas.

Fonte: Os Autores.

A média aritmética neste caso será dada por:

$$\bar{x} = \frac{87^\circ + 343^\circ + 71^\circ + 112^\circ + 86^\circ + \dots + 89^\circ}{11} \approx 176,82^\circ \quad (2.12)$$

o que não reflete adequadamente a situação, pois o valor encontrado de $176,82^\circ$ está bem distante da realidade plotada dos voos das abelhas apresentados na Figura (2.4) da esquerda. No entanto, com o uso da estatística circular, tem-se como resultado os seguintes valores:

1- Média dos cossenos obtida por (2.5):

$$\bar{C} = \frac{1}{11}(\cos 87^\circ + \cos 343^\circ + \cos 71^\circ + \dots + \cos 89^\circ) = 0,3532. \quad (2.13)$$

2- Média dos senos dada por (2.6):

$$\bar{S} = \frac{1}{11}(\sin 87^\circ + \sin 343^\circ + \sin 71^\circ + \dots + \sin 89^\circ) = 0,4762. \quad (2.14)$$

3- Média angular de (2.11):

$$\bar{\theta} = \arctg\left(\frac{0,4762}{0,3532}\right) = 53,4382^\circ. \quad (2.15)$$

4- Comprimento do vetor médio resultante em (2.8):

$$\bar{R} = \sqrt{(0,3532)^2 + (0,4762)^2} \approx 0,5929. \quad (2.16)$$

Nota-se neste exemplo que a média direcional ou angular do voo das 11 rainhas é de aproximadamente $53,44^\circ$, graficamente esta representado à direita da Figura(2.4), e uma possível análise preliminar desses dados com apenas esta parte teórica apresentada, tem-se que existe uma certa dispersão das direções, pois o comprimento médio do vetor resultante foi de $\bar{R} = 0,5929$, um valor que exprime relativamente uma heterogeneidade das direções dos voos.

2.2 Mediana Circular

A mediana direcional β dos ângulos $\theta_1, \theta_2, \theta_3, \dots, \theta_n$ é definida como um ângulo Φ que satisfaz as seguintes condições [5]:

1. A metade dos dados estejam no arco $[\Phi, \Phi + \pi)$;
2. A maioria dos dados são mais próximos de Φ do que $\Phi + \pi$.

Quando o espaço amostral n é ímpar, a direção da mediana é um dos dados, ou seja, será um único ângulo, mais precisamente, o dado da $\frac{n+1}{2}$ – ésima posição. Quando n é par a direção da mediana é a média aritmética angular entre os dois ângulos centrais do espaço amostral.

Em relação ao exemplo das rainhas da Tabela (2.1), se a análise da mediana for feita através da estatística linear, tem-se que o conjunto de dados possuem 11 elementos, ou seja, n é ímpar e, ordenando-os por ordem crescente, obtém-se como resultado o valor que ocupa a 6^a posição desta amostra. De fato,

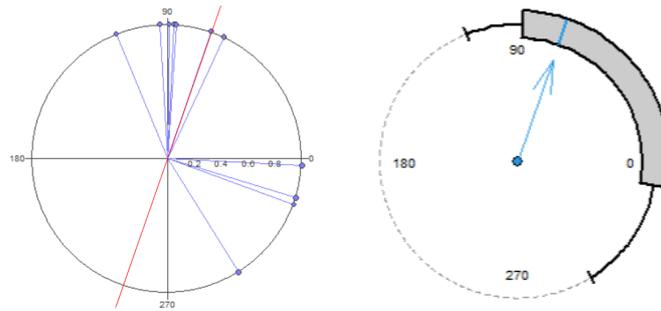
$$(65^\circ \ 71^\circ \ 86^\circ \ 87^\circ \ 89^\circ \ 93^\circ \ 112^\circ \ 302^\circ \ 340^\circ \ 343^\circ \ 357^\circ)$$

Portanto, a mediana (Me) é igual a 93° . O que não condiz com a direção esperada considerando a Figura (2.4). Em contrapartida, embasado na estatística circular, a direção da mediana Φ tem que satisfazer as seguintes exigências, das quais a metade dos dados estejam no arco $[\Phi, \Phi + \pi)$ e a maioria dos dados estejam mais próximos de Φ do que $\Phi + \pi$.

Deste modo, a mediana direcional será $\Phi = 71^\circ$, pois, a metade dos dados estão contidos no intervalo $[71^\circ, 251^\circ)$ e a maioria dos dados estão mais próximos de 71° do que do ângulo 251° . Graficamente, será representado um gráfico de pontos angulares das rainhas mas

também o denominado Boxplot circular, sendo este último um gráfico que tem como característica transmitir a idéia de posição, dispersão e assimetria dos dados, bem como a posição da mediana.

Figura 2.5: Mediana direcional do primeiro voo, em graus, das 11 rainhas, representada pelo gráfico dos pontos angulares e pelo Boxplot circular.



Fonte: Os autores.

2.3 Medidas de Dispersão

A **variância circular** amostral (V) é uma medida de dispersão circular definida por:

$$V = 1 - \bar{R}; \quad 0 \leq V \leq 1. \quad (2.17)$$

Neste sentido, quanto menor for o valor de V , maior será a concentração dos dados. Análise análoga a variância utilizada em dados lineares[1]. Novamente, considerando o exemplo das abelhas (2.1), percebe-se que a estatística usual não fornece dados confiáveis, observe que ao calcular a variância de forma linear tem-se um valor bem distante dos dados estudados.

$$s^2 = \sum_{i=1}^{11} \frac{(x_i - \bar{x})^2}{10} = \frac{(87 - 176,82)^2 + \dots + (357 - 176,82)^2}{10} \approx 16133,56. \quad (2.18)$$

De acordo com a definição (2.17) de variância para dados angulares e utilizando do resultado do comprimento do vetor resultante em (2.16), tem-se o seguinte:

$$V = 1 - \bar{R} \Rightarrow V = 1 - 0,5929 = 0,4071 \quad (2.19)$$

Segundo [5], paralelamente a estatística para dados lineares, há a necessidade de se definir o desvio padrão circular por uma transformação da variância V . Deste modo,

define-se como **Desvio Padrão Circular** a seguinte relação:

$$v = [-2 \ln(1 - V)]^{\frac{1}{2}} \Leftrightarrow v = [-2 \ln(\bar{R})]^{\frac{1}{2}} \in [0, \infty) \quad (2.20)$$

Para finalizar a análise entre aplicar a estatística descritiva linear em vez da estatística circular e, obter valores em certos estudos equivocados, como visto até aqui a partir do Exemplo (2.1), segue abaixo os cálculos dos seguintes desvios.

Para a estatística linear tem-se:

$$s = \sqrt{s^2} \Rightarrow s = \sqrt{(16133,56)} = 127,02^\circ \quad (2.21)$$

Com base na definição de desvio padrão angular (2.20), observa-se o seguinte:

$$v = [-2 \ln(1 - V)]^{\frac{1}{2}} \Rightarrow v = \sqrt{-2 \ln(1 - 0,4071)} \approx 1,0225 \text{ rad} \approx 58,59^\circ \quad (2.22)$$

Desta maneira, para fins de comparação e análise dos dados até aqui calculados, pode-se ver, conforme Tabela (2.2) a seguir, que os dados angulares e lineares das respectivas medidas analisadas são bem distintas.

Tabela 2.2: Comparação dos resultados do exemplo do voo das 11 abelhas rainhas obtidos através da estatística linear e angular.

Estatística	Média	Mediana	Desvio padrão
Angular	53,44°	71°	58,59°
Linear	176,82°	93°	127,02°

Fonte: Os Autores.

Portanto, verifica-se que quando se quer estudar dados angulares é necessário efetuar um estudo mais cauteloso e embasado na estatística descritiva circular, visto que a estatística usual fornece valores totalmente distintos dos valores reais e bem distantes da interpretação exigida em um estudo estatístico.

2.4 Distribuição de Probabilidade Circular

As propriedades das funções de distribuição circular e as funções de densidade de probabilidade possuem muitos aspectos em comum com as funções que são discutidas na estatística linear [7]. No entanto, há também importantes diferenças que decorrem na necessidade de se definir as distribuições advindas de dados circulares.

Uma distribuição circular é uma distribuição de probabilidade cuja probabilidade total está concentrada na circunferência em um círculo unitário. Uma vez que cada ponto da circunferência representa uma direção, tal distribuição é uma forma de atribuir probabilidades para direções diferentes ou definindo uma distribuição direcional.[8, pg-38]

Em especial, as funções de densidade de probabilidade $f(\lambda)$, também conhecidas pela sigla (fdp), existem se, e só se, as seguintes propriedades forem satisfeitas.

1. $f(\lambda) \geq 0$;
2. $f(\lambda + 2k\pi) = f(\lambda)$ para $k \in \mathbb{Z}$;
3. $\int_0^{2\pi} f(\lambda) d\lambda = 1$.

As distribuições circulares, essencialmente, as distribuições que relacionam as variáveis contínuas e as massas de probabilidades são chamadas de distribuições contínuas.

Vale destacar que nas distribuições contínuas existem vários modelos probabilísticos circulares, dos quais pode-se destacar a distribuição uniforme, a von Mises, Cardioide, Arqueadas (Wrapped) entre outras. Neste tópico, será discutido, de forma bem sucinta, dentre os vários modelos probabilísticos circulares, a distribuição circular uniforme conhecida como a mais simples e a distribuição de von Mises, dita como a mais importante. Para melhores detalhes, consultar [5], [8].

2.4.1 Distribuição Uniforme Circular

A função de densidade de probabilidade uniforme circular é definida como:

$$f(\lambda) = \frac{1}{2\pi}; 0 \leq \lambda \leq 2\pi. \quad (2.23)$$

A distribuição uniforme é dita como a mais básica no círculo, é a única distribuição no círculo que é invariante sob rotação e reflexão. Haja vista que na distribuição uniforme todas as direções são igualmente prováveis, ou seja, não existe uma direção preferencial e além disso o comprimento do vetor resultante é sempre igual a zero [5].

2.4.2 Distribuição de von Mises

Diante da semelhança com a distribuição normal (distribuição de Gauss) da estatística linear, a distribuição de von Mises é considerada como uma das mais importantes distribuições de probabilidade circular, visto que muitas de suas propriedades analisadas, tais

como, simetria, parâmetros e a própria regularidade são examinadas de forma análoga à estatística linear, assim, a distribuição de von Mises - $vM(\bar{\theta}, \kappa)$, uma função a dois parâmetros, será definida como a seguinte função de densidade de probabilidade:

$$f(\lambda; \bar{\theta}, \kappa) = \frac{1}{2\pi \mathbf{I}_0(\kappa)} e^{\kappa \cos(\lambda - \bar{\theta})} \quad (2.24)$$

onde $\bar{\theta}$ é a direção média, $0 \leq \bar{\theta} < 2\pi$ e κ é o parâmetro de concentração, $\kappa \geq 0$. Tem-se ainda, que $\mathbf{I}_0(\kappa)$ representa a função de Bessel ¹ modificada de primeiro tipo e ordem zero, dada pela seguinte igualdade:

$$\mathbf{I}_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa \cos(\lambda)} d\lambda = \sum_{r=0}^{\infty} \left(\frac{\kappa}{2}\right)^{2r} \left(\frac{1}{r!}\right)^2 \quad (2.25)$$

Uma observação importante em relação ao parâmetro de concentração κ e a função (2.24) é que quando $\kappa \rightarrow 0$, a função densidade de von Mises se aproxima da distribuição circular uniforme, ou seja, quando $\kappa \rightarrow 0$ as direções dos dados tendem a uma probabilidade igualmente provável na circunferência, por outro lado, quando $\kappa \rightarrow \infty$ menor será a variabilidade, apresentando então uma alta concentração dos dados [8].

2.5 Testes de hipóteses

Semelhante a estatística linear a estatística circular possui alguns testes de hipóteses, como por exemplo, o teste de Rao, teste de Rayleigh, teste de Watson, teste de Kuiper. A principal finalidade desses testes é estabelecer a qualidade de ajustamento dentre as distribuições de probabilidade já mencionadas neste trabalho [9].

Mas, especificamente neste trabalho, o teste que será evidenciado será o teste de Rayleigh. O teste de Rayleigh é usado para verificar se uma população está distribuída uniformemente em torno da circunferência. A hipótese nula (H_0) afirma que a população apresenta uma distribuição circular uniforme, já a hipótese alternativa (H_1), por sua vez, afirma que a população não é uniformemente distribuída em torno da circunferência e, portanto, existe uma direção preferida. O cálculo aplicado neste teste tem como base um valor Z cuja sua determinação é dada pela seguinte expressão:

$$Z = n\bar{R}^2, \text{ onde } n \text{ é o número de observações dos dados} \quad (2.26)$$

¹A função de Bessel foi definida primeiramente por Daniel Bernoulli, sendo obtida a partir da solução de uma equação diferencial ordinária de segunda ordem.

Sabe-se que o \bar{R} de certa forma, transmite a idéia de concentração dos dados. Quando a amostra segue uma distribuição uniforme circular, normalmente o valor do \bar{R} será pequeno e quando o valor do vetor médio resultante assume valores próximos a 1, a hipótese nula é rejeitada. Assim, quanto maior for o vetor médio resultante, maior será o valor de Z (2.26), o que significa uma maior concentração dos dados em torno da média, resultando em uma menor probabilidade destes dados estarem uniformemente distribuídos [9].

A equação que determina o cálculo da probabilidade Z é descrita pela seguinte igualdade:

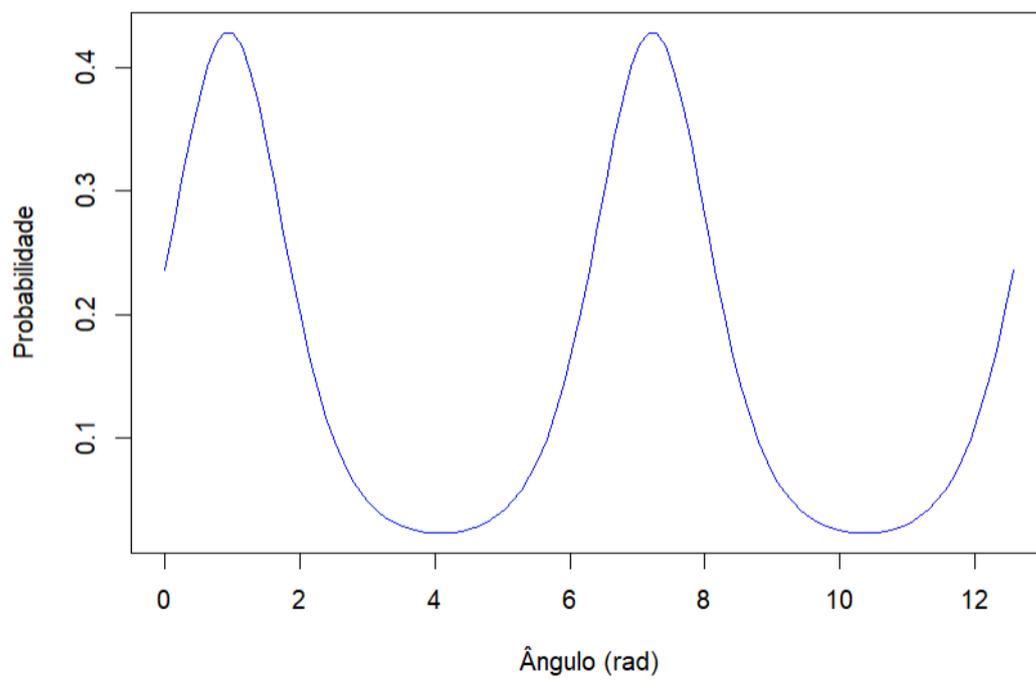
$$P(Z \geq z) = e^{\left[\sqrt{1+4n+4(n^2-nz)}-(1+2n)\right]} \quad (2.27)$$

Normalmente o nível de significância escolhido é 0,05, com isso uma probabilidade menor à este valor indica que os dados não estão distribuídos uniformemente, evidenciando uma direção tendenciosa dos dados.

Considerando os resultados obtidos da Tabela (2.1) relacionados ao exemplo do voo nupcial das 11 rainhas, embasados na estatística circular, tem-se que de acordo com o teste de Rayleigh o valor Z calculado é de aproximadamente 3,87 e além disso a probabilidade calculada, segundo equação (2.27) resulta em um valor-p igual a 0,0172, um valor inferior ao nível de significância adotado de 0,05, portanto, o voo das rainhas não estão distribuídos uniformemente em torno do círculo. Logo, a distribuição de probabilidade que modela este exemplo é von Mises(0,9326; 1,4805).

Em relação a curva de densidade von Mises das 11 abelhas rainhas (2.6) juntamente com o boxplot circular (2.5), ao analisar o gráfico de forma semelhante as realizadas na curva de distribuição normal gaussiana, nota-se uma assimetria em torno das medidas de posição, $\bar{\theta} = 0,9326 \text{ rad} \approx 53,44^\circ$ e $\Phi = 1,2391 \text{ rad} = 71^\circ$, que representam, principalmente, a direção média dos voos nupciais das abelhas. Mas, percebe-se ainda, que existe uma leve dispersão dos dados, pois o valor do desvio padrão calculado foi de ($v = 1,0225 \text{ rad} = 58,59^\circ$) o que reafirma o que foi citado nas conclusões de (2.16) relacionadas ao vetor médio resultante ($\bar{R} = 0,5929$) do exemplo das abelhas.

Figura 2.6: Gráfico da função densidade de probabilidade von Mises das 11 rainhas.



Fonte: Os Autores.

3 USO DA ESTATÍSTICA CIRCULAR PARA DADOS GEOGRÁFICOS

Dados angulares podem ser obtidos de dados geográficos. Segundo Falconi [10], para se obter as distâncias entre dois pontos geográficos, é necessário que se obtenha a diferença entre as medidas de latitude (DLA) e as diferenças de longitude (DLO) das coordenadas geográficas desses pontos, a partir daí, é preciso que os valores angulares obtidos da DLA e DLO sejam transformados em distâncias. O valor das distâncias relacionadas as diferenças de latitude e de longitude são definidas pela seguinte expressão:

$$\text{DLA ou DLO} = \left[g(60) + \text{min} + \frac{s}{60} \right] (1.852) \text{ metros.} \quad (3.1)$$

tal que as constantes presentes g , min e s são as medidas de graus, minutos e segundos, respectivamente. É importante frisar que 1 minuto da circunferência terrestre equivale a uma milha náutica (NM), e uma milha náutica equivale a 1.852 metros. Com o uso das relações trigonométricas entre as localidades, o ângulo (a), Figura (3.1), será encontrado da seguinte maneira:

$$\text{tg}(a) = \frac{\text{DLA}}{\text{DLO}} \Rightarrow a = \text{arctg} \left(\frac{\text{DLA}}{\text{DLO}} \right) \quad (3.2)$$

A partir deste princípio, podem ser calculados os ângulos e distâncias entre duas cidades de interesse. Na cidade de Barbacena, MG, está localizada a Escola Preparatória de Cadetes do Ar (EPCAR) que é uma organização de ensino do comando da Aeronáutica cuja missão é preparar seus alunos para o ingresso no Curso de Formação de Oficiais Aviadores (CFOAV) na Academia da Força Aérea (AFA). A EPCAR é uma escola de ensino médio militar, na qual os estudantes aprovados tem acesso a uma formação completa, combinada com instrução militar e física. A admissão é feita por meio de concurso público, que inclui provas escritas, exames médicos e testes físicos. Durante os três anos de estudo, os alunos recebem uma educação de alta qualidade, que inclui as disciplinas acadêmicas

comuns exigidas no ensino médio, além de uma educação militar que abrange tópicos de disciplina, ética, liderança e trabalho em equipe.

Neste trabalho, serão abordados dados reais relacionados ao perfil dos alunos ingressantes no Curso Preparatório de Cadetes do Ar (CPCAR) desta instituição. Os dados refere-se aos estados de origem dos alunos aprovados nas turmas denominadas Anhur, Hércules e Cavok, 3^o, 2^o e 1^o ano respectivamente. Este conjunto de dados será utilizado para verificar a dispersão dos ingressantes nesta escola, haja vista, que este estabelecimento possui um concurso à nível federal e anual, as provas são aplicadas em todos os estados do país e para que se tenha uma noção da relação candidato/vaga dessas três turmas, a média geral entre as inscrições deferidas foram de 126, 75 e 135 por vaga dos ingressantes em 2020, 2021 e 2022 respectivamente, ou seja, um concurso muito disputado. Diante disso, a cada ano, a busca ao êxito da aprovação se torna ainda mais desafiador, sem contar que os ingressantes na EPCAR, possuem moradia, planos de saúde, um plano de carreira e ainda são remunerados com uma bolsa mensalmente, cujo valor, atualmente, esta em torno de 1044 reais.

Inicialmente, foi realizado o levantamento dos estados de origem de cada aluno dos três esquadrões e assumiu-se como coordenadas geográficas de localização, latitude (LA) e longitude (LO), as coordenadas relacionadas as suas respectivas capitais disponibilizadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Para determinar as distâncias e as direções angulares, foi fixado a cidade de Barbacena/MG, sede da EPCAR, como ponto referencial, ou seja, Barbacena foi considerada como o ponto de origem de um sistema de eixos cartesianos, e as demais capitais como sendo as coordenadas geográficas de origem dos estados que possuem ao menos um aprovado no concurso. Para exemplificar essa situação, será exposto os cálculos dos alunos oriundos do estado de Minas Gerais, considerando as coordenadas geográficas da capital Belo Horizonte, conforme representação gráfica a seguir.

De fato, analisando o exemplo da Figura (3.1) tem-se os seguintes resultados.

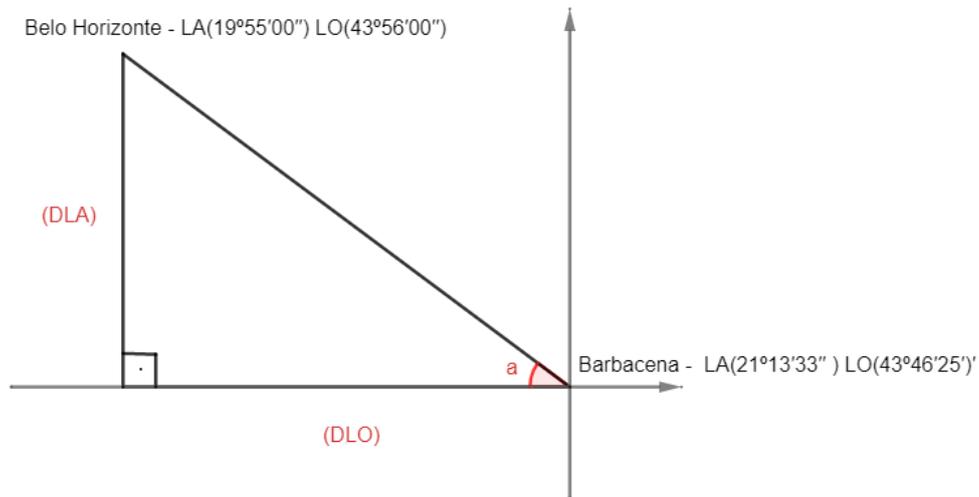
- Barbacena

$$LA = 21^{\circ}13'33'' = 1.273,55 \text{ min}$$

$$LO = 43^{\circ}46'25'' \approx 2.626,42 \text{ min}$$

- Belo Horizonte

Figura 3.1: Coordenadas geográficas de Barbacena e Belo Horizonte em um sistema de eixos cartesianos



Fonte: Os Autores.

$$LA = 19^{\circ}55'00'' = 1.195 \text{ min}$$

$$LO = 43^{\circ}56'00'' = 2.636 \text{ min}$$

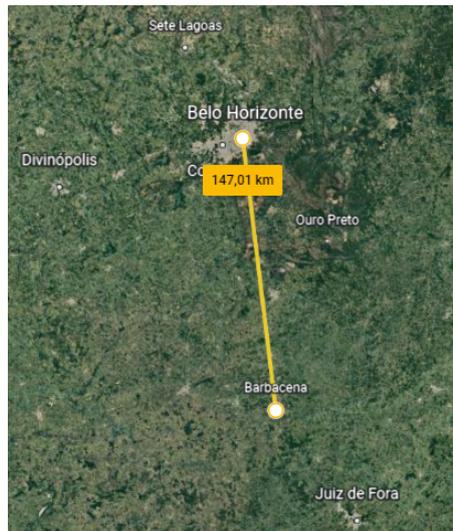
- $DLA = 1.273,55 - 1.195 = 78,55 \Rightarrow (78,55)(1.852) = 145.474,6$
- $DLO = 2.626,42 - 2.636 = -9,58 \Rightarrow (-9,58)(1.852) = -17.742,16$

$$\text{tg}(a) = \frac{DLA}{DLO} = \frac{145.474,6}{-17.742,16} \Rightarrow a = \text{arctg} \left(\frac{145.474,6}{-17.742,16} \right) \approx -83,0465$$

De (2.11), pode-se ver que quando $DLA > 0$ e $DLO < 0$ o ângulo pertencerá ao segundo quadrante, deste modo, é necessário que seja adicionado 180° para ajustar o quadrante, assim, $a = -83,0465 \approx 97^{\circ}$. Portanto, segue que a direção angular de Belo horizonte em relação a cidade de Barbacena será de 97° . Para o cálculo de distância entre as localidades, leva-se em consideração a distância em linha reta das cidades. Assim, com suporte nas relações trigonométricas, ao aplicar o teorema de Pitágoras, segue o valor da distância entre Barbacena e Belo Horizonte:

$$\begin{aligned} d^2 &= DLA^2 + DLO^2 \\ &= (145.474,6)^2 + (-17.742,16)^2 \Rightarrow \\ d &\approx 147.000 \text{ metros} \approx 147 \text{ km} \end{aligned}$$

Figura 3.2: Distância em linha reta entre Barbacena e Belo Horizonte



Fonte: Os Autores.

Repetindo estes procedimentos para as demais localidades, juntamente com o auxílio do software R [11] e seus pacotes Circular e CircStats em busca de suporte matemático para a exibição dos resultados, foi calculado os demais ângulos e distâncias entre Barbacena e as capitais dos vários estados contemplados com pelo menos um aluno aprovado no CPCAR, como se pode observar nas Tabelas (3.1), (3.3) e (3.5) listadas a seguir. Posteriormente a isso, serão apresentados resultados relacionados as medidas de posição e dispersão embasados de acordo com a teoria da estatística circular e linear a fins de análise e comparação entre resultados obtidos, conforme descritos nas Tabelas (3.2), (3.4) e (3.6). E também serão listados os seguintes gráficos, denominados como gráfico de pontos, diagrama de rosas e os boxplot angular e linear respectivamente.

Por meio da Tabela (3.1) referente a turma Anhur, ou seja, o 3º ano, segue os resultados relacionados às medidas de posição e dispersão, segundo as teorias citadas de estatística, conforme Tabela (3.2).

Tabela 3.1: Estados de origem dos alunos do 3º esquadrão da EPCAR, ingressantes no ano de 2020 e suas respectivas coordenadas geográficas considerando as distâncias das capitais em relação à Barbacena - MG.

Estado	Número de alunos	Latitude	Longitude	Direção(°)	Distância(km)
Barbacena/MG	-	21°13'33"	43°46'25"	0	0
CE	1	3°43'06"	38°32'34"	73,4	2031
DF	6	15°47'38"	47°52'58"	127,1	757
ES	1	20°19'08"	40°20'16"	14,8	395
GO	1	16°40'00"	49°15'00"	140,3	792
MG	27	19°55'00"	43°56'00"	97,0	147
MS	2	20°28'53"	54°36'58"	176,1	1208
MT	1	15°35'45"	56°05'49"	155,5	1506
PA	1	1°27'21"	48°30'14"	103,5	2259
PE	6	8°03'14"	34°52'51"	56,0	1766
PR	5	25°25'47"	49°16'19"	217,4	769
RJ	82	22°54'10"	43°12'28"	288,7	197
RN	6	5°47'42"	35°12'32"	61,0	1961
RS	3	30°01'58"	51°13'48"	229,8	1283
SC	2	27°35'49"	48°32'56"	233,2	885
SP	24	23°33'01"	46°38'02"	219,1	410
Total	168	-	-	-	

Fonte: Os Autores

Tabela 3.2: Parâmetros dos dados circulares e lineares dos alunos do 3º esquadrão da EPCAR, ingressantes no ano de 2020.

Medidas Angulares			Medidas Lineares
parâmetros	radianos	graus	valores
$\bar{\theta}$	-1,621028	267,2	$\bar{x} = 214,9^\circ$
\bar{R}	0,3251445	-	-
Φ	5,038766	288,7	Me = 231,5°
V	0,6748555		$s^2 = 7473,93$
v	1,49899	85,9	$s = 86,45^\circ$
κ	0,6876911	-	-
Teste de Rayleigh (valor-p)	< 0,05		

Fonte: Os Autores

Para a turma do 2º ano tem-se os demais dados relacionados a origem e a distância dos alunos ingressantes na EPCAR no ano de 2021, turma Hércules e, a partir dos cálculos angulares e lineares desta turma, serão listados os resultados da turma na Tabela (3.4).

Tabela 3.3: Estados de origem dos alunos do 2º esquadrão da EPCAR, ingressantes no ano de 2021 e suas respectivas coordenadas geográficas considerando as distâncias das capitais em relação à Barbacena - MG.

Estado	Número de alunos	Latitude	Longitude	Direção (°)	Distância(km)
Barbacena/MG	-	21°13'33"	43°46'25"	0	0
AL	1	9°39'57"	35°44'06"	55,2	1565
BA	2	12°58'13"	38°30'45"	57,5	1088
CE	3	3°43'06"	38°32'34"	73,4	2031
DF	5	15°47'38"	47°52'58"	127,1	757
ES	1	20°19'08"	40°20'16"	14,8	395
GO	3	16°40'00"	49°15'00"	140,3	792
MG	25	19°55'00"	43°56'00"	97,0	147
MS	2	20°28'53"	54°36'58"	176,1	1208
PB	4	7°07'08"	34°52'53"	57,8	1853
PE	5	8°03'14"	34°52'51"	56,0	1766
PI	1	5°05'20"	42°48'07"	86,6	1797
PR	4	25°25'47"	49°16'19"	217,4	769
RJ	108	22°54'10"	43°12'28"	288,7	197
RN	8	5°47'42"	35°12'32"	61,0	1961
RO	1	8°45'43"	63°54'14"	148,2	2631
RR	1	2°49'12"	60°40'19"	132,6	2777
RS	5	30°01'58"	51°13'48"	229,8	1283
SC	2	27°35'49"	48°32'56"	233,2	885
SE	1	10°54'36"	37°04'12"	57,0	1367
SP	33	23°33'01"	46°38'02"	219,1	410
Total	215	-	-	-	

Fonte: Os Autores

Tabela 3.4: Parâmetros dos dados circulares e lineares dos alunos do 2º esquadron da EPCAR, ingressantes no ano de 2021.

Medidas Angulares			Medidas Lineares
parâmetros	radianos	graus	valores
$\bar{\theta}$	-1,539578	271,8	$\bar{x} = 216,5^\circ$
\bar{R}	0,3575781	-	-
Φ	5,038766	288,7	Me = 288,7°
V	0,6424219		$s^2 = 7654,23$
v	1,434156	82,2	$s = 87,49^\circ$
κ	0,7657484	-	-
Teste de Rayleigh (valor-p)	< 0,05		

Fonte: Os Autores

Analogamente as turmas do terceiro e segundo ano, segue a Tabela (3.6) com os resultados estudados da turma Cavok em relação aos ângulos e distâncias das demais localidades que contemplam algum aprovado nesta turma.

Tabela 3.5: Estados de origem dos alunos do 1º esquadrão da EPCAR, ingressantes no ano de 2022 e suas respectivas coordenadas geográficas considerando as distâncias das capitais em relação à Barbacena - MG.

Estado	Número de alunos	Latitude	Longitude	Direção(°)	Distância(km)
Barbacena/MG	-	21°13'33"	43°46'25"	0	0
AM	1	3°4'25"	60°0'20"	130,7	2662
BA	4	12°58'13"	38°30'45"	57,5	1088
CE	1	3°43'06"	38°32'34"	73,4	2031
DF	4	15°47'38"	47°52'58"	127,1	757
ES	2	20°19'08"	40°20'16"	14,8	395
GO	3	16°40'00"	49°15'00"	140,3	792
MG	22	19°55'00"	43°56'00"	97,0	147
MS	1	20°28'53"	54°36'58"	176,1	1208
PA	2	1°27'21"	48°30'14"	103,5	2259
PB	1	7°07'08"	34°52'53"	57,8	1853
PE	1	8°03'14"	34°52'51"	56,0	1766
PR	2	25°25'47"	49°16'19"	217,4	769
RJ	50	22°54'10"	43°12'28"	288,7	197
RN	4	5°47'42"	35°12'32"	61,0	1961
RO	1	8°45'43"	63°54'14"	148,2	2631
RR	2	2°49'12"	60°40'19"	132,6	2777
RS	5	30°01'58"	51°13'48"	229,8	1283
SE	1	10°54'36"	37°04'12"	57,0	1367
SP	19	23°33'01"	46°38'02"	219,1	410
Total	126	-	-	-	

Fonte: Os Autores

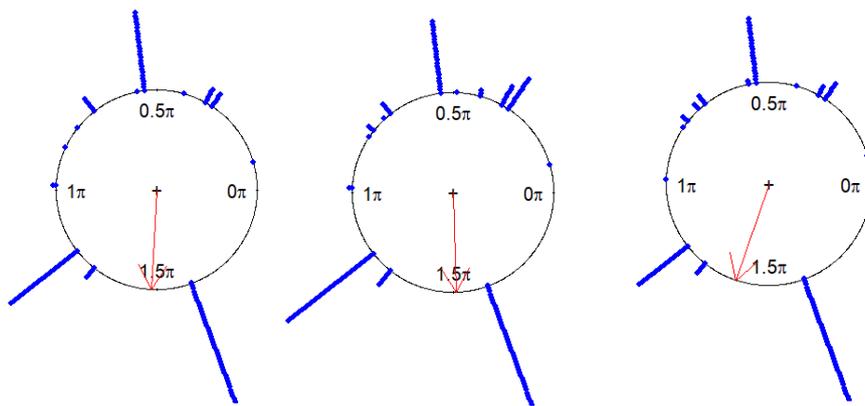
Tabela 3.6: Parâmetros dos dados circulares e lineares dos alunos do 1º esquadrão da EPCAR, ingressantes no ano de 2022.

parâmetros	Medidas Angulares		Medidas Lineares
	radianos	graus	valores
$\bar{\theta}$	-1,894139	251,5	$\bar{x} = 197,8^\circ$
\bar{R}	0,1825491	-	-
Φ	5,038766	288,7	Me = 219,1°
V	0,8174509		$s^2 = 8083,16$
v	1,844308	105,7	$s = 89,91^\circ$
κ	0,3713504	-	
Teste de Rayleigh (valor-p)	< 0,05		

Fonte: Os Autores

Em seguida, estão listados os demais gráficos de pontos e o diagrama de rosas das respectivas turmas, Anhur, Hércules e Cavok, que apresentam em sua essência, as frequências das vagas ocupadas em relação as direções angulares. E a fim de comparação entre a estatística circular e a estatística linear, estão listados os gráficos denominados boxplot angular e linear, uma análise que se nota entre estes gráficos é o fato que se o estudo fosse embasado pela estatística linear, os dados apresentariam uma dispersão muito mais alta que de fato ocorre na distribuição das vagas, ou seja, o boxplot linear aparenta uma dispersão das vagas muito maior que a realidade, sem contar que nesta análise não existem nenhum ponto discrepante nos dados das turmas Hércules e Cavok, o que não é verdade segundo os boxplot circulares dos respectivos anos.

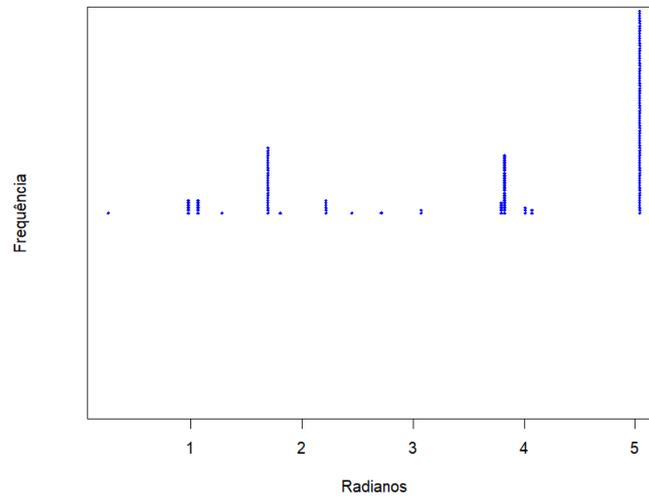
Figura 3.3: Gráfico de pontos das turmas ingressantes nos anos de 2020, 2021 e 2022 respectivamente.



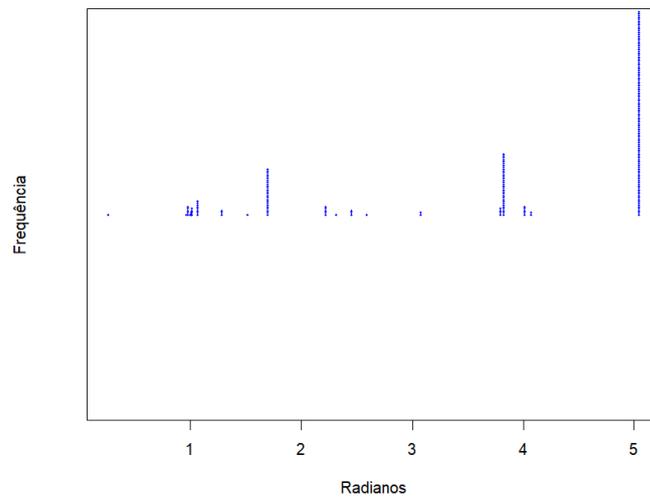
Fonte: Os Autores.

Figura 3.4: Diagrama de pontos das turmas ingressantes nos anos de 2020, 2021 e 2022 respectivamente.

Ingressantes em 2020 - 3^o ano.



Ingressantes em 2021 - 2^o ano.



Ingressantes em 2022 - 1^o ano.

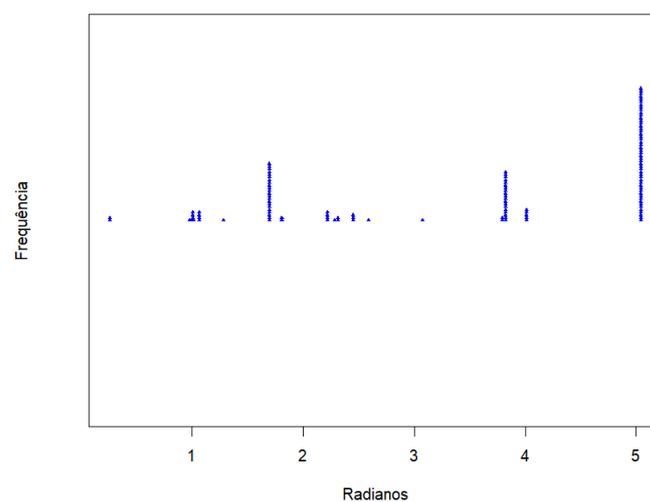
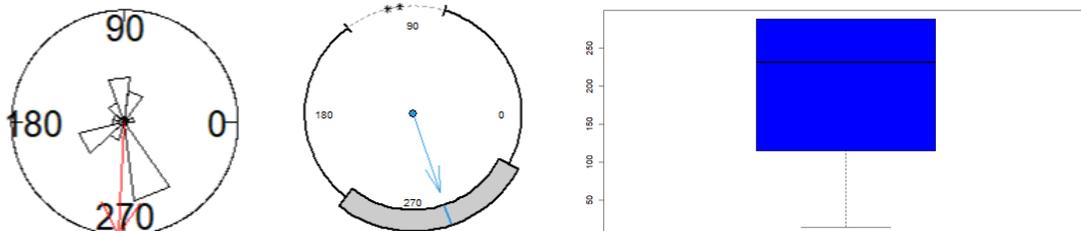


Figura 3.5: Diagrama de rosas, boxplot circular e linear da turma ingressante no ano de 2020.

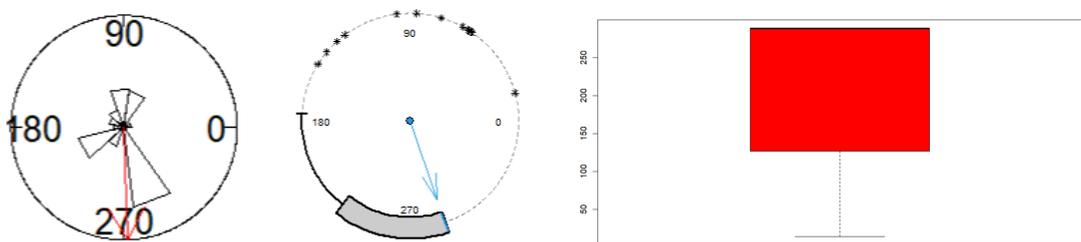
Ingressantes em 2020 - 3^o ano.



Fonte: Os Autores.

Figura 3.6: Diagrama de rosas, boxplot circular e linear da turma ingressante no ano de 2021.

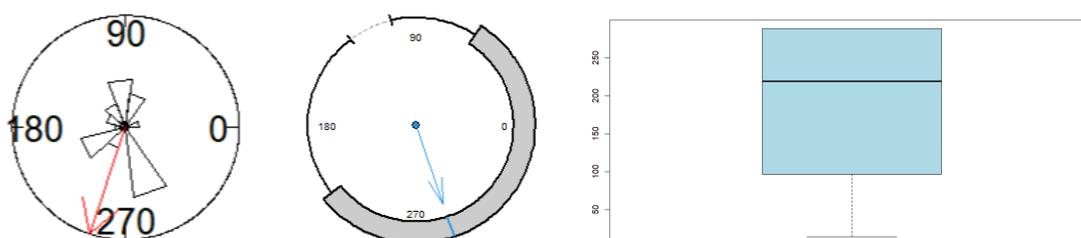
Ingressantes em 2021 - 2^o ano.



Fonte: Os Autores.

Figura 3.7: Diagrama de rosas, boxplot circular e linear da turma ingressante no ano de 2022.

Ingressantes em 2022 - 1^o ano.



Fonte: Os Autores.

4 DISCUSSÕES

Para que se possa ter condições de argumentação em relação ao ingresso dos alunos no concurso do CPCAR nos diferentes anos analisados até aqui, foi aplicado o teste de Rayleigh afim de verificar se de fato os dados analisados possuem uma distribuição uniforme em torno de uma circunferência. A partir deste teste, verificou-se a rejeição da hipótese nula (H_0), pois, segundo os resultados do valor-p, destacados nas Tabelas (3.2), (3.4), (3.6), todos são inferiores ao nível de significância adotado de 0,05, conseqüentemente, considerou-se a hipótese alternativa (H_1) que afirma que os dados possuem uma tendência direcional, o que nos dá suporte para um melhor ajuste das distribuições de probabilidade, ou seja, os dados possuem uma distribuição von Mises.

Diante dos fatos, foi realizado o ajuste dos dados conforme as definições de distribuição de probabilidade de von Mises, o que confirmou uma periodicidade direcional em relação aos alunos aprovados no CPCAR. Percebe-se claramente no Gráfico (4.1) que nos anos de 2020, 2021 e 2022, ingressantes avaliados neste trabalho, tanto a média direcional quanto a mediana para dados circulares são valores acima de 4 e próximos a 5 radianos, algo em torno de 4,7 rad aproximadamente, valores que remetem a direção angular do estado do Rio de Janeiro. O que confirma os cálculos apresentados nas Tabelas (3.2), (3.4) e (3.6) do terceiro, segundo e primeiro ano, pois, os valores das médias direcionais calculados foram $\bar{\theta} = -1,621028 \text{ rad} = 267,2^\circ$, $\bar{\theta} = -1,539578 \text{ rad} = 271,8^\circ$ e $\bar{\theta} = -1,894139 \text{ rad} = 251,5^\circ$ respectivamente, observa-se ainda, uma predominância das vagas ocupadas no concurso da região sudeste, algo em torno de 70% e 80%, valores muito significativos e dominantes.

Um possível destaque que vale a pena enfatizar neste momento, é o fato que neste período de 2020 a 2022 ocorreu a pandemia causada pelo vírus SARS-CoV-2 (COVID-19). O vírus foi identificado pela primeira vez em Wuhan, na China, em 31 de dezembro de 2019, e desde então se espalhou globalmente, causando uma pandemia que afetou milhões de pessoas em todo o mundo [12].

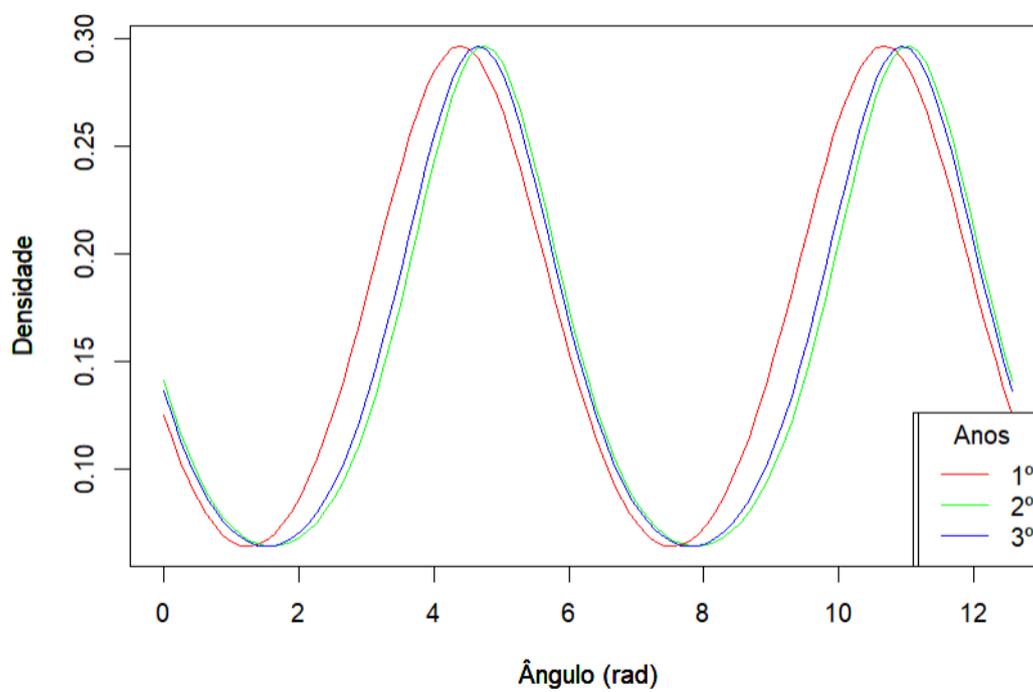
Com base nisso, o terceiro ano, turma Anhur incorporada em 2020, realizou o exame de admissão em um período sem pandemia, em condições normais, visto que a prova foi realizada em 07/07/2019.

Já a turma Hércules ingressante em 2021, realizou a prova no auge da pandemia, de fato, em 11/03/2020 a Organização Mundial de Saúde (OMS) declarou o início da pandemia e em 20/03/2020 o Ministério da Saúde declara distanciamento social, o que impactou totalmente a rotina de estudos nas escolas, ou seja, foi necessário a implementação de modelos de aulas até então distantes da rotina e com realidades bem diferentes em relação aos anos anteriores, observa-se ainda que, normalmente, as provas da EPCAR são aplicadas nas últimas semanas de junho ou nas primeiras semanas de julho, porém neste período o Brasil já registrava um total de 64365 mortes, o que ocasionou adiamento do exame para 18/10/2020.

Em relação ao primeiro ano, turma Cavok ingressante em 2022, esta turma fez a prova do concurso em 04/07/2021, ano em que já havia uma esperança de controle da doença devido o início da campanha de vacinação em 17/01/2021, mas mesmo assim, a pandemia ainda estava causando muitas mortes, por exemplo em 26/04/2021 o número de óbitos já ultrapassava o total de mortes do ano de 2020 e um pouco antes do dia do exame, o Brasil já se aproximava de 510 mil mortes pela doença.

Assim, com base nessas informações e principalmente nos resultados dos boxplots circulares (3.5), (3.6) e (3.7) relacionados as demais turmas, tem-se que estes gráficos além de apresentarem informações relacionadas a existência de assimetria nas vagas ocupadas, mostram uma concentração regional nas distribuições dos ingressantes em relação ao país. O principal destaque vai para a turma que ingressou em 2021, em se tratando de um concurso nacional, nesta turma a pandemia influenciou negativamente no preenchimento das vagas disponibilizadas, as vagas preenchidas ficaram ainda mais concentradas e com muito mais pontos discrepantes. Porém a turma que entrou em 2022, também na pandemia, apresentou uma maior dispersão, o que nos remete em uma entrada que contemplou em alunos oriundos de outros estados que até então não possui um alto índice de aprovação. Mas mesmo assim, nota-se que os exames realizados em anos com contextos nacionais e sanitários totalmente distintos, ainda sim manteve-se a periodicidade e a média direcional deste concurso nacional.

Figura 4.1: Gráfico da função densidade de probabilidade von Mises das 3 turmas.



Fonte: Os Autores.

5 CONSIDERAÇÕES FINAIS

Embasados na estatística circular viu-se que seus métodos, até aqui citados, permitiram que os dados estudados, oriundos de coordenadas geográficas fossem avaliados de forma consistente, o que a estatística linear não proporcionou, além do mais foi possível uma avaliação sobre a concentração dos dados bem como sua distribuição de probabilidade.

Verificou-se através do teste de Rayleigh uma tendência direcional dos alunos que são aprovados no concurso, confirmando a rejeição da hipótese nula que garantia a aleatoriedade dos aprovados.

A ideia deste trabalho foi apresentar, principalmente as medidas de tendência central, levando em consideração a má aplicação da teoria da estatística linear em dados cuja abordagem envolvem dados periódicos, angulares ou direcionais. Além do mais, foi citado, afins de conhecimento, diferentes ferramentas de distribuição de probabilidade e testes de hipóteses, claro, dando ênfase e apresentando a parte teórica da distribuição de von Mises e do teste de Rayleigh. O que mostrou serem ferramentas essenciais na análise direcional dos alunos ingressantes na EPCAR.

Obteve-se também a apresentação de gráficos que até então não são citados no ensino básico, como por exemplo o diagrama de rosas e o boxplot circular, o que com suas apresentações podem vir a estimular um olhar crítico dos alunos do ensino médio para dados angulares, ou seja, é possível mostrar aos alunos, a partir de um simples exemplo, cuja organização de dados se dá em torno de uma circunferência, tal que cada ponto desses dados esteja localizado em um ângulo proporcional ao seu valor e até mesmo uma breve introdução das medidas de posição que envolvem cálculos trigonométricos, ou seja, apresentar uma breve introdução conceitual sobre a importância em se ter técnicas coerentes diante de dados que possuam direções. No caso, por exemplo, da EPCAR, o uso poderia ser inclusive para análise de resultados em competições de tiro ao alvo e/ou dardos sobre as conclusões em relação aos conceitos de precisão, viés e acurácia.

No mais, uma reflexão diante dos aprovados e da própria instituição sobre a necessidade de melhorar a divulgação nas regiões com menos incidência de aprovados, ou até mesmo incluir políticas de acesso a fim de amenizar o regionalismo das vagas ocupadas.

REFERÊNCIAS

- 1 SILVA, C. T. S. da; BRIGHENTI, C. R. G.; RESENDE, L. F. S. Estatística circular aplicada aos dados de localização dos municípios de origem dos alunos do profmat-ufsj-campus santo antonio. *Revista de Matemática*, v. 1, n. 01, p. 1–10, 2022.
- 2 MENEZES, E. B. Um estudo direcional dos estudantes que chegam à UFRB: Uma aplicação da estatística circular. 2018.
- 3 BARBOSA, L. M.; QUEIROZ, P. I. B. de; JÚNIOR, W. C. de S. Estatística de ventos e análise de condições de dispersão atmosférica em São José dos Campos/SP. NINFA informe técnico n. 03. São José dos Campos: CTA/ITA, 2021. Disponível em: <<https://redevale.ita.br>>. Acesso em: 23/04/2023.
- 4 BRASIL. Ministério da Educação. Base Nacional Comum Curricular. Brasília, DF, 2018. Disponível em: <<http://basenacionalcomum.mec.gov.br/#/site/inicio>>. Acesso em: 15/09/2022.
- 5 MARDIA, K. V. Statistics of directional data. In: . [S.l.: s.n.], 1972.
- 6 AYRES, M. et al. Bioestat: aplicações estatísticas nas áreas das ciências biomédicas. *Ong Mamiraua. Belém, PA*, 2007.
- 7 PEWSEY, A.; NEUHÄUSER, M.; RUXTON, G. D. *Circular statistics in R*. [S.l.]: Oxford University Press, 2013.
- 8 JAMMALAMADAKA, S. R.; SENGUPTA, A. *Topics in circular statistics*. [S.l.]: world scientific, 2001. v. 5.
- 9 GÓIS, J. E. S. Contribuição dos modelos estocásticos para o estudo da climatologia urbana. 2002.
- 10 FALCONI, C. E. *Calculando distâncias e direções utilizando coordenadas geográficas*. 2009. Disponível em: <<https://www.pilotopolicial.com.br/calculando-distancias-e-direcoesutilizando-coordenadas-geograficas/>><https://www.pilotopolicial.com.br/calculando-distancias-e-direcoesutilizando-coordenadas-geograficas/>. Acesso em: 03 nov 2022.
- 11 R., R. D. C. T. *R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria, 2021. ISBN 3-900051-07-0*. 2021. Disponível em: <<https://www.r-project.org/>><https://www.r-project.org/>. Acesso em: 03 ago 2022.
- 12 SANAR. *Linha do Tempo do Coronavírus no Brasil*. Disponível em: <<https://www.sanarmed.com/linha-do-tempo-do-coronavirus-no-brasil>>. 2020. Acesso em: 23 abr 2023.

APÊNDICE A

Comandos utilizados no software R

```
aluno <- read.delim("C:/Users/Elton/Desktop/aluno3.txt")
```

```
aluno
```

```
library(circular)
```

```
library(CircStats)
```

```
library(NPCirc)
```

```
library(bpDir)
```

```
library(ggplot2)
```

```
tab=rep(aluno$graus,aluno$ni)
```

```
tab
```

```
alunorad=rad(tab)
```

```
alunorad
```

```
mean.circular(alunorad)
```

```
###comprimento do vetor###
```

```
rho.circular(alunorad)
```

```
###mediana direcional###
```

```
median.circular(alunorad)
```

```
var.circular(alunorad)
```

```
###desvio padrão para dados angulares ###
```

```
sd.circular(alunorad)
```

```
###Parâmetro de concentração###
```

```
est.kappa(alunorad)

###plotar gráficos(Diagrama de rosas)###

rose.diag(alunorad, bins=13, prop=1, pts=F, cex= 0.9, pch=10, dotsep=40, shrink=
1.0)

arrows.circular(mean.circular(alunorad), zero = 0, rotation = "counter", col = "red")

###Gráficos com pontos:###

plot.circular(cbind(alunorad), rotation = "counter", bins = 80, zero = 0, stack = T,
units = "rads", axes = T, col = c("black", "grey"), ticks = F) ###boxplot###

circ= as.circular(alunorad)

CircularBoxplot (circ, template="degrees", place="none", units="radians", marg=
"large", shrink = 1.5, H=F, stack=F, constant= "optimal")

###Estatística linear###

media= mean(tab)

media

mediana= median(tab)

mediana

variancia= var(tab)

variancia

desviopadrao= sqrt(variancia)

desviopadrao

boxplot(tab, col = "blue")

Diagrama de pontos

stripchart(alunorad, vertical=F, xlab="Radianos", col= "blue", ylab="Frequência", off-
set=0.26, pch = 18, method="stack", frame= T, cex= 0.5)

### von mises ###

aluno1 <-read.delim("C:/Users/Elton/Desktop/aluno.txt")

aluno1
```

```
tab1=rep(aluno1$graus,aluno1$ni)
tab1
alunorad1=rad(tab1)
alunorad1
circ.disp(alunorad1)
circ.mean(alunorad1)
circ.summary(alunorad1)
est.kappa(alunorad1)
aluno2 <- read.delim("C:/Users/Elton/Desktop/aluno2.txt")
aluno2
tab2=rep(aluno2$graus,aluno2$ni)
tab2
alunorad2=rad(tab2)
alunorad2
circ.disp(alunorad2)
circ.mean(alunorad2)
circ.summary(alunorad2)
est.kappa(alunorad2)
aluno3 <- read.delim("C:/Users/Elton/Desktop/aluno3.txt")
aluno3
tab3=rep(aluno3$graus,aluno3$ni)
tab3
alunorad3=rad(tab3)
alunorad3
circ.disp(alunorad3)
circ.mean(alunorad3)
```

```
circ.summary(alunorad3)

est.kappa(alunorad3)

plot(function(x) dvm(x,-1.894139, 0.3713504), 0, 4*pi, axes= F, col= "red", main="Densidade
de Distribuição Von Mises relacionada aos três esquadros",xlab = "ângulo (rad)", ylab=
"Densidade")

vm=function(x) dvm(-1.894139, 0.1825491, 0.3713504)

par(new= T)

plot(function(x) dvm(x,-1.539578, 0.7657484), 0, 4*pi, col= "green", xlab = "ângulo
(rad)", ylab= "Densidade")

vm=function(x) dvm(-1.539578, 0.3575781, 0.7657484, add= T)

par(new= T)

plot(function(x) dvm(x,-1.621028, 0.6876911), 0, 4*pi, axes= F, col= "blue", xlab =
"ângulo (rad)", ylab= "Densidade")

vm=function(x) dvm(-1.621028, 0.3251445,0.6876911)

legend("bottomright", inset= 0.0001, c("1º", "2º", "3º"), cex= 1,lty=1,
col=c("red", "green", "blue"),title="Anos")

### Teste de Rayleigh###

rayleigh.test(alunorad)
```