
Estatística circular aplicada aos dados de localização dos municípios de origem dos alunos do PROFMAT-UFSJ - Campus Santo Antonio

Clodoaldo Teodosio Santana da Silva

teoelania@gmail.com

Universidade Federal dos Vales do Jequitinhonha e Mucuri, Teófilo Otoni, MG, Brazil

Universidade Federal de Lavras, Lavras, MG, Brasil

Carla Regina Guimarães Brighenti

carlabrighenti@ufsj.edu.br

Universidade Federal de São João del Rei, São João del Rei, MG, Brazil

Luiz Fernando Silva Resende

luiz.resende@estudante.ufla.br

Universidade Federal de Lavras, Lavras, MG, Brazil

Resumo

O PROFMAT - Programa de Mestrado Profissional em Matemática em Rede Nacional - é um programa criado com o objetivo de contribuir com a qualificação dos professores da educação básica. O curso é ofertado em Rede Nacional sob coordenação da SBM (Sociedade Brasileira de Matemática) e o IMPA (Instituto de Matemática Pura e Aplicada) em parceria com 77 instituições de ensino superior em todo o Brasil, entre elas a Universidade Federal de São João del Rei - MG (UFSJ). Os alunos matriculados no curso são oriundos, em sua maioria, de municípios próximos à São João del Rei. Assim, estabelecer a abrangência, ou mesmo a distribuição de localização de origem dos alunos atendidos pelo programa no entorno da instituição de ensino superior parceira da SBM, pode ser um fator de interesse para gestão de coordenadores e, posterior divulgação entre os professores de matemática. Desta forma, ter acesso as coordenadas geográficas dos municípios de origem dos alunos fornece um conjunto de dados de interesse, sendo que tais dados podem ser transformados em dados angulares em relação à cidade sede do curso. Para realizar análises com dados angulares se faz necessário o uso da estatística circular, isto é, a análise considerando dados que não devem ser organizados simplesmente na reta real e sim estabelecendo uma disposição circular. Diante do exposto, o objetivo deste trabalho foi utilizar a estatística circular para analisar a distribuição de localização de origem dos alunos matriculados no PROFMAT na UFSJ, Campus Santo Antonio.

Palavras-chave

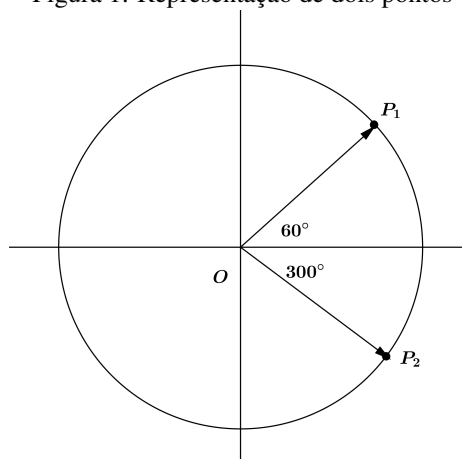
Média circular, Variância circular, Comprimento médio.

1 Introdução

Quando se pretende estudar fenômenos que são provenientes de medidas que envolvem ângulos como, direção dos ventos, ângulos de voos de pássaros, a melhor ferramenta é a estatística circular. No entanto muitas técnicas estatísticas são mais adequadas para dados localizados na reta real, como altura, idade e peso. Estes dados são considerados como lineares.

Contudo os dados circulares não devem ser analisados da mesma maneira como se tratam os lineares. O motivo desta afirmação está no fato de que se medidas angulares fossem tratadas como lineares seriam obtidos resultados equivocados. Por exemplo, a média aritmética dos ângulos 60° e 300° é 180° , no entanto observando a Figura 1, nota-se que o valor mais apropriado para esta média é 0° . Nota-se então que há vantagem de utilizar a estatística circular para que se obtenha a interpretação correta dos dados. Neste contexto os dados analisados são medidas angulares, ou que podem ser

Figura 1: Representação de dois pontos



Fonte: Os autores

transformadas em ângulos, como é o caso dos meses do ano e coordenadas geográficas. Já existem vários estudos utilizando dados circulares, pode ser mencionado o uso da estatística circular para analisar o efeito da direção dos ventos e a concentração de ozônio [3] e um outro estudo sobre a dispersão dos alunos matriculados na Universidade Federal do Recôncavo da Bahia- UFRB [7].

O objetivo desta pesquisa foi analisar, via estatística circular, a localização das cidades de origem dos alunos do programa de mestrado, PROFMAT, da UFSJ - Campus Santo Antônio, a partir das coordenadas geográficas. A relevância deste estudo está no fato de que é possível verificar a distribuição dos alunos nas cidades situadas na região de atuação da UFSJ.

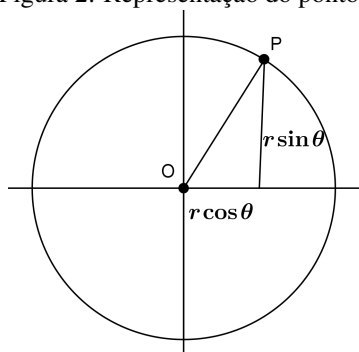
Isto tornará possível verificar em que localidades é mais importante divulgar o curso, tendo em vista a necessidade de qualificação dos professores da educação básica.

Para realização da pesquisa foram analisadas as 26 cidades mineiras de origem dos 56 alunos matriculados no PROFMAT no período de 2017- 2021.

2 Estatística Circular Descritiva

Um ponto P do plano cartesiano pode ser representado por (x, y) , este ponto também pode ser visto como um ponto numa circunferência de centro na origem $(0, 0)$ e raio r , assim pode ser estabelecida uma relação : $x = r \cos \theta$ e $y = r \sin \theta$, então $P = (r \cos \theta, r \sin \theta)$ (Figura 2).

Figura 2: Representação do ponto P



Fonte: Os autores

No caso em que o raio é unitário, isto é, $r = 1$, tem-se $x = \cos \theta$ e $y = \sin \theta$, um ponto P num círculo unitário pode ser representado por $P = (\cos \theta, \sin \theta)$.

A partir deste fato será dado o conceito da **Média direcional**, a qual tem a mesma direção que o centro de massa do círculo unitário (ver [6]).

Assim, considere agora os pontos P_1, P_2, \dots, P_n em uma circunferência unitária, e seus respectivos ângulos $\theta_1, \dots, \theta_n$ em relação ao eixo x . O centro de massa será dado por $M = (\bar{C}, \bar{S})$ onde

$$\bar{C} = \frac{1}{n} \sum_{i=1}^n \cos \theta_i \tag{1}$$

$$\bar{S} = \frac{1}{n} \sum_{i=1}^n \sin \theta_i \tag{2}$$

Seja

$$\bar{R}^2 = \bar{C}^2 + \bar{S}^2 \tag{3}$$

Note que das equações (1) e (2) tem-se que $C = \sum_{i=1}^n \cos \theta_i$ e $S = \sum_{i=1}^n \sin \theta_i$

A **média direcional**, $\bar{\theta}$, é a solução das equações :

$$\begin{cases} \cos \bar{\theta} = \frac{\bar{C}}{\bar{R}} \\ \sin \bar{\theta} = \frac{\bar{S}}{\bar{R}} \end{cases} \quad (4)$$

e satisfaz as equações (5)

$$\bar{\theta} = \begin{cases} \arctan \frac{\bar{S}}{\bar{C}}, S > 0, C > 0 \\ \arctan \frac{\bar{S}}{\bar{C}} + \pi, C < 0 \\ \arctan \frac{\bar{S}}{\bar{C}} + 2\pi, S < 0, C > 0 \end{cases} \quad (5)$$

A medida \bar{R} é chamada de **comprimento médio resultante**, observe que $\bar{R} = \frac{R}{n}$, com $0 \leq \bar{R} \leq 1$, e quanto mais próximo de 1 está o valor de \bar{R} menos dispersos estarão os dados. A partir deste fato, pode-se obter uma medida para avaliar a dispersão dos dados. Se \bar{R} tem um valor próximo de 1, então os dados direcionais estão bem agrupados, caso \bar{R} esteja próximo de zero há uma dispersão dos dados. Notando assim, que \bar{R} é uma medida de concentração dos dados([5],[2]).

A **variância circular** V é dada por

$$V = 1 - \bar{R} \quad (6)$$

onde $0 \leq V \leq 1$.

Observe que quanto mais próximo de 0(zero) for o valor na variância circular mais concentrados estão os dados e quanto mais próximo de 1 há uma dispersão dos dados.

O **desvio padrão circular** é dado por $\nu = [-2 \ln(1 - V)]^{\frac{1}{2}}$ ([6])

Pode também ser utilizado uma aproximação para o desvio padrão, ν , que é dada por

$$\nu \approx (2V)^{\frac{1}{2}} \quad (7)$$

Considerando os valores de θ tais que $0 \leq \theta \leq 1$ rad, e usando a expansão de

Taylor (com apenas dois termos) para $\cos \theta$ (ver [4]) tem-se:

$$\cos \theta \approx 1 - \frac{\theta^2}{2} \tag{8}$$

De (8)

$$2 \cos \theta \approx 2 - \theta^2$$

$$\theta^2 \approx 2(1 - \cos \theta)$$

$$\sum_{i=1}^n \theta_i^2 \approx 2 \sum_{i=1}^n (1 - \cos \theta_i)$$

Com isso

$$\sum_{i=1}^n (\theta_i - \bar{\theta})^2 \approx 2 \sum_{i=1}^n (1 - \cos(\theta_i - \bar{\theta})) \tag{9}$$

Antes de desenvolver a igualdade (9), será mostrado que

$$\sum_{i=1}^n \sin(\theta_i - \bar{\theta}) = 0 \tag{10}$$

$$\sum_{i=1}^n \cos(\theta_i - \bar{\theta}) = R \tag{11}$$

Usando relações trigonométricas e as equações (4) tem-se:

$$\begin{aligned} \sum_{i=1}^n \sin(\theta_i - \bar{\theta}) &= \cos \bar{\theta} \sum_{i=1}^n \sin \theta_i - \sin \bar{\theta} \sum_{i=1}^n \cos \theta_i \\ &= S \frac{\bar{C}}{\bar{R}} - C \frac{\bar{S}}{\bar{R}} \\ &= S \frac{C}{R} - C \frac{S}{R} = 0 \end{aligned}$$

e

$$\begin{aligned} \sum_{i=1}^n \cos(\theta_i - \bar{\theta}) &= \sum_{i=1}^n (\cos \theta_i \cos \bar{\theta} + \sin \theta_i \sin \bar{\theta}) \\ &= \cos \bar{\theta} \sum_{i=1}^n \cos \theta_i + \sin \bar{\theta} \sum_{i=1}^n \sin \theta_i \\ &= C \cos \bar{\theta} + S \sin \bar{\theta} = \frac{C^2}{R} + \frac{S^2}{R} = R \end{aligned}$$

Retornando à equação (9) :

$$\begin{aligned} \sum_{i=1}^n (\theta_i - \bar{\theta})^2 &\approx 2 \sum_{i=1}^n (1 - \cos(\theta_i - \bar{\theta})) \\ \sum_{i=1}^n (\theta_i - \bar{\theta})^2 &\approx 2n - 2 \sum_{i=1}^n \cos(\theta_i - \bar{\theta}) = \\ &= 2(n - R) = 2(n - n\bar{R}) \\ &= 2n(1 - \bar{R}) \end{aligned}$$

Então

$$\begin{aligned} \sum_{i=1}^n (\theta_i - \bar{\theta})^2 &\approx 2n(1 - \bar{R}) \\ \frac{\sum_{i=1}^n (\theta_i - \bar{\theta})^2}{n} &\approx 2(1 - \bar{R}) \\ \nu^2 &\approx 2(1 - \bar{R}) \end{aligned}$$

Portanto

$$\nu \approx \sqrt{2(1 - \bar{R})} = \sqrt{2V} \tag{12}$$

De acordo com [5] para valores da variância circular V próximos de zero, pode ser utilizada a aproximação dada em (12) para o desvio padrão circular ν .

3 Metodologia

Os dados foram obtidos através das informações dos estudantes do PROFMAT registradas no Sistema Integrado de Gestão das Atividades Acadêmicas da Universidade Federal de São João Del Rei, localizada no município de São João Del Rei – MG. Para este trabalho, fez-se uso dos dados provenientes das cidades de origem dos estudantes matriculados no curso ofertado no campus Santo Antônio. Foi analisado o período de 2017 – 2021, no qual há 56 alunos matriculados oriundos de 26 cidades mineiras.

Utilizou-se métodos estatísticos circulares para averiguar as frequências direcionais dos dados fornecidos. A partir das coordenadas geográficas de cada município fornecida pelo Instituto Brasileiro de Geografia Estatística (IBGE), foram obtidas as distâncias e direções para determinação das medidas angulares, tendo a cidade sede da UFSJ como município de referência.

Para o cálculo das distâncias, de acordo com [1] se faz necessário obter a distância entre dois pontos geográficos P_1 e P_2 se calculando então a diferença de latitude (DLA) e diferença de longitude (DLO). Para se obter os valores da DLA e DLO é necessário calcular a diferença entre as latitudes e longitudes entre as coordenadas de P_1 e P_2 . Essas diferenças são aplicáveis para pequenas distancias, segundo [1] para ângulos

maiores que 7 graus deve-se levar em consideração a curvatura da terra. Calculada as diferenças, o próximo passo é transformar os valores angulares em distância, que generalizamos como D, porém para o cálculo das diferença de latitude e diferença de longitude chamaremos respectivamente de DLA e DLO.

$$D = [G(60) + M + \frac{S}{60}](1852) \quad (13)$$

A medida D está quantificada em graus(G), minutos(M) e segundos(S), o valor 1852 é referente a transformação em Milhas Náuticas. Por meio da expressão (13) as diferenças (D) foram convertidas em distância, considerando que uma milha náutica (1 NM), equivale 1 minuto da circunferência terrestre. Assim tem-se que uma milha equivale a 1852 metros ou 1,852Km.

Pelas propriedades de um triangulo retângulo, sabe-se que a tangente é a divisão do cateto oposto pelo adjacente, seja P_1 a origem em um plano cartesiano, ou seja, o ponto (0, 0), podemos deduzir que o ângulo α do seguimento P_1 e P_2 em um círculo trigonométrico, será:

$$\tan(\alpha) = \frac{DLA}{DLO} \quad (14)$$

Sabendo-se o valor da tangente, o valor de α será dado por:

$$\arctan |(\alpha)| = \tan(\alpha^{-1}) \quad (15)$$

O valor de α é dado em modulo pois se faz necessário encontrar o quadrante ao qual α pertence, e o quadrante será dado de acordo com os sinais do DLA e DLO.

A cidade de São João del Rei – MG foi considerada como 0° (zero grau). Para realização das análises utilizou-se o pacote circular e CircStats do software R [8].

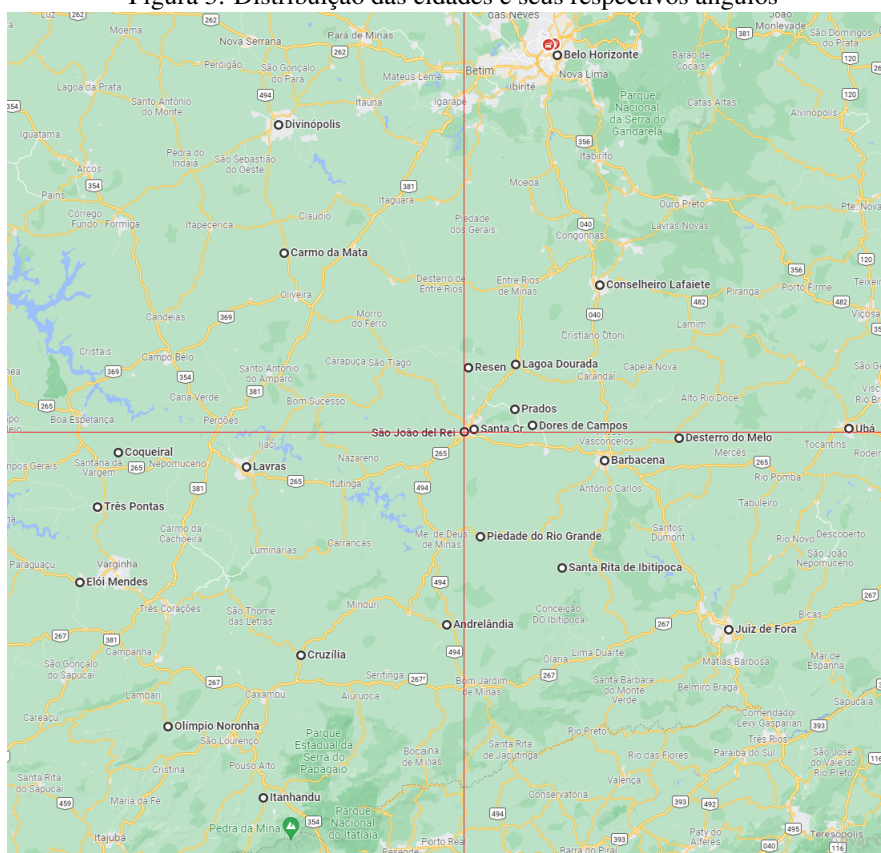
4 Resultados e Discussões

A Tabela 1 mostra os ângulos encontrados e a distância, em relação à cidade de referência, para cada uma das 26 cidades de origem dos alunos matriculados no PROFMAT.

A distribuição dos alunos, nas diversas cidades, está representada nas Figuras 3 e 4, e percebe-se que alunos de várias regiões do entorno de São João del Rei frequentam este programa de pós graduação.

Considerando que os dados estão distribuídos numa circunferência de raio 1, a

Figura 3: Distribuição das cidades e seus respectivos ângulos



Source: Os autores

média circular obtida foi $\bar{\theta} = 0,398$ rad ou $\bar{\theta} = 22^\circ$, o comprimento médio resultante foi de $\bar{R} = 0,29$ e a variância circular $V = 0,71$. Note que os alunos estão distribuídos em toda a região do entorno de São João del Rei. Além disso, os alunos estão mais concentrados nas cidade situadas ao norte da cidade sede do curso.

Em seguida foi encontrada a média circular e a variâncias dos dados

5 Conclusão

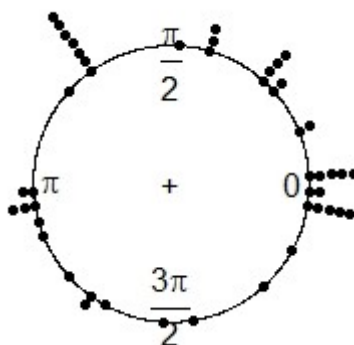
Os dados dos alunos do programa de pós graduação PROFMAT indicam que a procura de alunos pelo curso é proveniente das diversas regiões nas quais a UFSJ está inserida. Estes dados podem ser um indicativo da necessidade de formação de professores em nível mestrado profissional em Matemática nestas regiões, e ainda sugerir a necessidade de adequações na universidade e na cidade de São João de Rei, no sentido da melhoria na infraestrutura para a recepção destes alunos.

Não há uma concentração dos alunos matriculados no curso em apenas uma região

Cidade	Núm. de alunos	Lat.	Longit.	Direção (°)	Distância(Km)
São João del Rei	7	21° 8' 9"	44° 15' 43"	0	0,00
Santa Cruz de Minas	1	21° 7' 12"	44° 13' 22"	22	4,69
Ubá	2	21° 7' 12"	42° 56' 34"	1	146,60
Prados	1	21° 3' 28"	44° 4' 48"	23	22,00
Resende Costa	1	20° 55' 19"	44° 14' 16"	84	23,54
Lagoa Dourada	3	20° 54' 50"	44° 4' 48"	50	32,05
Dores de Campos	1	21° 6' 32"	44° 1' 22"	6	26,70
Barbacena	9	21° 13' 33"	43° 46' 26"	350	55,14
Desterro de Melo	2	21° 8' 49"	43° 31' 4"	359	82,7
Ouro Branco	1	20° 31' 15"	43° 41' 31"	47	93,17
Divinópolis	7	20° 8' 33, 6"	44° 53' 25, 2"	127	130,59
Conselheiro Lafaiete	2	20° 39' 36"	43° 47' 9"	45	74,80
Belo Horizonte	3	19° 55' 0"	43° 56' 0"	75	140,30
Carmo da Mata	1	20° 33' 28"	44° 52' 15"	44	93,30
Coqueiral	2	21° 11' 20"	45° 26' 27"	183	131,13
Lavras	2	21° 14' 42"	45° 0' 0"	188	82,90
Três Pontas	1	21° 22' 13"	45° 30' 44"	191	141,30
Varginha	1	21° 23' 3"	45° 25' 48"	199	137,70
Elói Mendes	1	21° 36' 36"	45° 33' 54"	200	154,00
Olímpio Noronha	1	22° 4' 4"	45° 15' 50"	223	152,00
Cruzília	2	21° 50' 20"	44° 48' 28"	232	99,00
Itanhandu	1	22° 17' 45"	45° 44' 56"	240	149,00
Andrelândia	1	21° 44' 24"	44° 18' 32"	266	67,30
Piedade do Rio Grande	1	21° 28' 8"	44° 11' 45"	281	37,7
Santa Rita de Ibitipoca	1	21° 33' 46"	43° 54' 54"	309	61,00
Juiz de Fora	1	21° 41' 20"	43° 20' 40"	329	119,00

Tabela 1: Cidades de origem dos alunos do PROFMAT entre 2017 e 2021 suas respectivas coordenadas geográficas com distâncias em relação à São João del Rei - MG

Figura 4: Gráfico circular com a distribuição dos alunos



Fonte: Os autores

específica. Cerca de 30% dos alunos são oriundos da região ao norte de São João del Rei. Com isso, há a indicativo de uma necessidade maior divulgação do curso nas cidades do entorno, o que poderia resultar numa demanda maior de professores da educação básica em se matricular no programa. Esta pesquisa pode ser estendida para outros cursos da instituição, tanto de graduação como de pós-graduação.

6 Agradecimentos

Os autores agradecem o apoio dado pela Universidade Federal dos Vales do Jequitinhonha e Mucuri, Universidade Federal de São João del Rei, e ao Departamento de Estatística da Universidade Federal de Lavras.

Referências

- [1] Falconi, Carlos Eduardo. Calculando distâncias e direções utilizando coordenadas geográficas. < <https://www.pilotopolicial.com.br/calculando-distancias-e-direcoes-utilizando-coordenadas-geograficas/>>, 2009.
- [2] Fisher, Nicholas. *Statistical analysis of circular data*. Cambridge University Press, New York, 1993.
- [3] Jammalamadaka, Sreenivasa Rao and Lund, Ulrich J. The effect of wind direction on ozone levels: a case study. *Environment ecology statistics*, (13):287–298, 2006.
- [4] Jammalamadaka, Sreenivasa Rao and SenGupta, A. *Topics in circular statistics*. World Scientific, New Jersey, 2001.
- [5] Kantil, Mardia and Jupp, Peter. *Directional statistics*. Jonh Wiley & Sons, London, 2000.
- [6] Mardia, Kantil. *Statistics of directional data*. Academic Press, London, 1972.
- [7] Menezes, Eliane Barbosa. Um estudo direcional dos estudantes que chegam à UFRB: Uma aplicação da estatística circular. [trabalho de conclusão de curso -TCC]. *UFRB - Universidade Federal do Recôncavo da Bahia*, 2018.
- [8] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. ISBN 3-900051-07-0.