

DISTRIBUIÇÃO VON MISES NA AVALIAÇÃO DE DADOS ENTOMOLÓGICOS

Yola Salomé Serpa de Miranda¹
Carla Regina Guimarães Brighenti²

¹Graduanda em Engenharia Química, Universidade Federal de São João Del Rei, São João Del Rei, MG.

² Prof. Dra. em Estatística e Experimentação Agropecuária, Universidade Federal de São João Del Rei – MG, Departamento de Engenharia de Biosistemas – DEPEB

Resumo: A estatística circular (ou direcional) é aplicada a dados dispostos em torno de uma circunferência. Medidas circulares, caracterizadas pelo horário do dia, são observadas em fenômenos periódicos como o ritmo circadiano do metabolismo de insetos, e este pode influenciar no horário de mortalidade desses em laboratório. O objetivo do trabalho foi verificar, através da estatística circular, a existência do ritmo circadiano na mortalidade em operárias de *Apis mellifera*. Para tanto utilizou-se os dados de dois experimentos laboratoriais nos quais a estatística circular foi aplicada. Conclui-se que há presença de ritmo circadiano no horário de mortalidade de *A. mellifera* que tendem a morrer durante o dia.

Palavras-chave: Estatística circular, ritmo circadiano, parâmetro de concentração

1 Introdução

Em diversas áreas do conhecimento, tais como Biologia, Meteorologia, Medicina e Engenharia, existe interesse por medidas na forma de ângulos denominadas na estatística como *dados circulares* ou *dados direcionais* (Leal & Souza, 2003).

As medidas circulares são registradas em fenômenos com observações direcionais (direção do movimento de um animal após um determinado estímulo, direção do vento, etc.) bem como em fenômenos periódicos (horário de chegada de um paciente a um hospital, visita de insetos às flores) (Izbicki & Esteves, 2008). Elas são obtidas a partir de uma origem arbitrária e as técnicas estatísticas desenvolvidas para variáveis reais não são, em geral, apropriadas para essas situações (Leal & Souza, 2003).

O acompanhamento da mortalidade de abelhas em experimentos laboratoriais depende de várias constantes. No entanto, a existência de um ritmo circadiano no metabolismo desses insetos pode influenciar no horário de mortalidade, sendo obtidas nesse caso, medidas circulares, caracterizadas pelo horário do dia. Withrow (1959) introduziu o termo circadiano (latim circa: em torno de; dies: dia), para caracterizar os ritmos com períodos de 24 horas, os quais são sincronizáveis em ciclos de 24 horas de claro/escuro. Segundo Harker (1958), citado por Proni & Macieira (2004), a característica de maior interesse no estudo de ritmo de 24 horas, não seria o fato de certas atividades ocorrerem todas neste intervalo, mas sim que tais repetições persistissem na ausência de mudanças ambientais normais.

1- Contato: yolaserpa@hotmail.com. Agradecimento à FAPEMIG pelo apoio financeiro

2- Contato: carlabrighenti@ufsj.edu.br

No caso bi-dimensional, uma observação de uma variável circular pode ser vista como um ponto P sobre o círculo unitário, centrado na origem O do sistema de coordenadas cartesianas e, portanto, tem-se que \overline{OP} é um vetor unitário em \mathbb{R}^2 . Essa observação circular pode ainda ser vista como um ângulo y , formado pelo vetor \overline{OP} e o semi-eixo O_x , medido no sentido anti-horário. Assim, uma observação de uma variável circular pode ser representada pelo par ordenado $(\cos(y), \text{sen}(y))$, em coordenadas cartesianas, ou pelo par $(1, y)$, em coordenadas polares (Leal & Souza, 2003).

Embora os dados circulares sejam coletados em diferentes tipos de unidades e possam ser expressos em qualquer medida de ângulo, adota-se trabalhar em radianos, com intervalo de variação $[0, 2\pi)$ ou $[-\pi, \pi)$, dependendo do interesse da análise (Izbicki & Esteves, 2008).

Para o estudo com distribuições de probabilidades referentes a variáveis aleatórias circulares é necessário estabelecer algumas medidas descritivas relacionadas com os conceitos de posição, dispersão, etc. (Leal & Souza, 2003).

Considerando-se uma amostra aleatória de n observações circulares y_1, \dots, y_n , às quais se pode associar n vetores unitários $\overline{OP}_1, \dots, \overline{OP}_n$. Uma maneira natural e intuitiva de resumir esses vetores unitários é através da sua soma. Do ponto de vista algébrico, calcula-se as seguintes quantidades (Izbicki & Esteves, 2008):

$$S = \sum_{i=1}^n \text{sen}(y_i), \quad C = \sum_{i=1}^n \cos(y_i), \quad R^2 = C^2 + S^2$$

A média circular de y_1, \dots, y_n é definida como sendo o ângulo $\bar{\mu}$ correspondente ao vetor resultante da soma $\overline{OP}_1, \dots, \overline{OP}_n$. Esse ângulo deve satisfazer as condições:

$$1) \cos(\bar{\mu}) = C/R$$

$$2) \text{sen}(\bar{\mu}) = S/R$$

$$3) \bar{\mu} = \begin{cases} \text{arctg}(S/C), & \text{se } S \geq 0 \text{ e } C > 0, \\ \text{arctg}(S/C) + \pi, & \text{se } C < 0, \\ \text{arctg}(S/C) + 2\pi, & \text{se } S < 0 \text{ e } C > 0. \end{cases}$$

A quantidade $R = \sqrt{C^2 + S^2}$ representa o comprimento do vetor resultante e fornece informação sobre o grau de concentração dos ângulos observados.

Em vez de R , é mais comum usar o comprimento médio do vetor resultante, definido por $\bar{R} = R/n$, o qual tem a vantagem de variar no intervalo $[0, 1]$. A variância circular amostral, ou parâmetro de concentração, é definida por $V = 1 - \bar{R}$, enquanto que o desvio padrão circular é definido por $\sigma = \{-2 \log(1 - V)\}^{1/2}$, diferentemente da definição usual para dados lineares, onde o desvio padrão corresponde diretamente à raiz quadrada da variância (Izbicki & Esteves, 2008).

A variância mínima ocorre quando $V = 0$ ($\bar{R} = 1$), e corresponde a todas as observações estarem precisamente no mesmo local. O maior limite possível de variação ocorre para dados uniformemente distribuídos ao redor do círculo, e corresponde a $V = 1$ ($\bar{R} = 0$). O cálculo de \bar{R} portanto, é simples e a interpretação dos resultados não depende de suposições sobre os dados originais (Otieno, 2002).

A mais importante distribuição para dados circulares é a chamada “Distribuição de von Mises”. Ela é tão importante para dados circulares, quanto o modelo Gaussiano é para dados lineares em geral. Esse modelo foi proposto pelo físico alemão R. von Mises, em 1918, ao investigar desvio de pesos atômicos de valores inteiros. Seu propósito era mostrar que os pesos atômicos são números inteiros sujeitos a erros. Transformando a parte fracionária dos pesos atômicos em desvios angulares, então, ele passou do problema físico original para um problema estatístico, a partir do qual originou essa distribuição (Mardia e Jupp, 2000).

Dizemos que uma variável aleatória circular Y tem distribuição von Mises com parâmetros μ e κ se sua função densidade de probabilidade for dada por:

$$f_Y(y; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(y - \mu)] \quad 0 \leq y < 2\pi, \quad 0 \leq \mu < 2\pi \quad \text{e} \quad \kappa > 0$$

onde $I_0(\kappa)$ é a função de Bessel modificada do 1º tipo e ordem zero, que é dada pela série

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

O parâmetro μ representa a média circular de Y , enquanto κ é denominado parâmetro de concentração.

O objetivo do trabalho foi verificar a presença do ritmo circadiano na mortalidade em operárias de *Apis mellifera*, utilizando medidas circulares.

2 Metodologia

A implementação computacional das operações, foi realizada utilizando-se os pacotes Circular e CircStat do software R Core Team (2009). A estatística circular do software foi aplicada a um conjunto de dados reais provenientes de dois experimentos realizados objetivando estudar o ritmo de mortalidade abelhas ao longo do dia. A estatística circular do software foi aplicada a dois conjuntos de dados reais provenientes de experimentos realizados objetivando estudar o ritmo de mortalidade de abelhas ao longo do dia. Para a diferenciação do horário de mortalidade foram realizados dois experimentos. Um dos experimentos foi montado durante o dia (7h30min) e o outro durante a noite (19h30min). O horário de mortalidade de cada grupo de 100 abelhas *A. mellifera* foi avaliado ao longo do dia.

Para a obtenção das medidas circulares de dispersão utilizou-se o comando circ.disp(x) em que x representa o vetor de observações medidas em radianos. A média direcional foi obtida através do comando circ.mean(x). O gráfico de dispersão circular foi obtido pelo comando circ.plot(x) e o diagrama do tipo “rose diagram” foi obtido pelo comando rose.diag(x). A estimativa da probabilidade máxima de kappa(κ), o parâmetro de concentração de uma distribuição von Mises, foi obtido pelo comando est.kappa(x).

4 Resultados e Discussões

As medidas circulares dos dados estão apresentadas na Tabela 1. Os dados foram testados em relação ao ajuste com duas distribuições: uniforme e von Mises, optando-se pela distribuição von Mises.

O gráfico de dispersão circular e o gráfico do tipo “rose diagram” estão representados respectivamente nas figuras 1 e 2.

Tabela 1 – Medidas de dispersão dos dados provenientes dos experimentos montados à noite e durante o dia

Horário de Montagem	n	$\bar{\mu}$ (rad)	\bar{R}	V	κ	$\bar{\mu}$ (horas)
Noite	82	-1.868201	0.894743	0.1052569	3.380006	10h22min
Dia	96	1.124270	0.820143	0.179857	2.516422	09h50min

n é o tamanho da amostra, $\bar{\mu}$ a média circular, \bar{R} é o comprimento médio resultante, V a variância e κ o parâmetro de concentração de uma distribuição von Mises

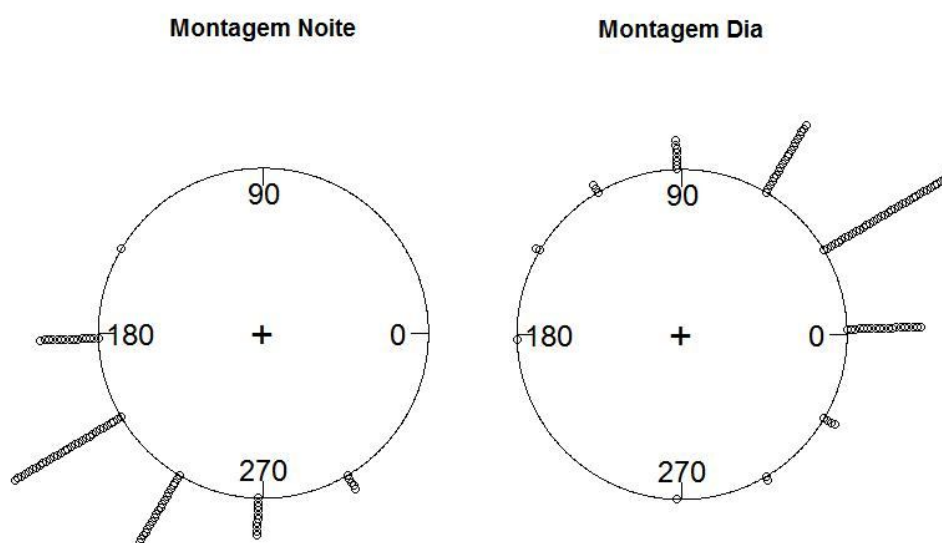


Figura 1 – Gráfico de dispersão circular para os experimentos montados à noite e de dia.

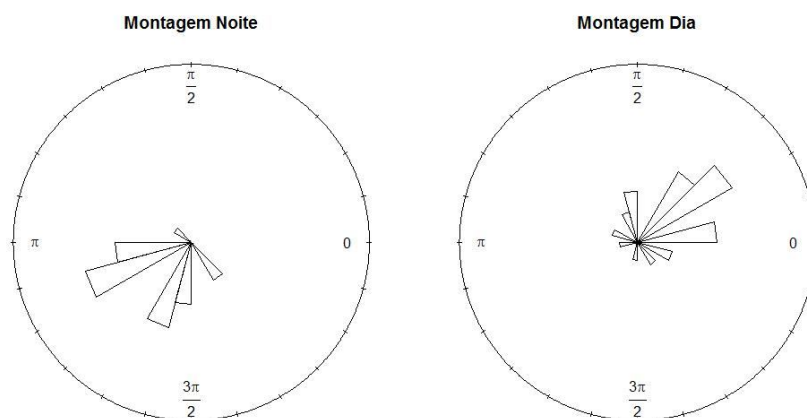


Figura 2 – Gráfico do tipo “rose diagram” para os experimentos montados à noite e de dia.

Observa-se nas figuras 1 e 2 que as mortes das abelhas se concentram no primeiro quadrante na montagem realizada durante o dia e no terceiro quadrante na montagem realizada durante a noite. Isso ocorre porque há uma diferença de 12 horas (180°), entre uma montagem e outra.

Como a média do horário de mortalidade nos dois casos é próxima (09h50min para a montagem realizada durante o dia e 10h22min para a montagem à noite), nota-se um “atraso” na mortalidade das abelhas quando a montagem do experimento é realizada durante o período noturno. Segundo Southwick (1992), citado por Brighenti et. al (2008), um inseto exposto a menor quantidade de luz tem menor atividade e, conseqüentemente, um possível aumento de sua longevidade. A consideração do horário de montagem do experimento é importante em experimentos de curta duração pois a diferença de horário de mortalidade pode sub ou superestimar os valores do tempo de vida ou de avaliação, por exemplo, de um inseticida.

5 Conclusões

1. A estatística circular se mostrou adequada pra avaliação de dados entomológicos.
2. Existe um ritmo circadiano que regula o horário de mortalidade das abelhas *A. mellifera*.
3. A taxa de mortalidade das abelhas *Apis mellifera* confinadas é afetada pelo horário da montagem do experimento.
4. Recomenda-se, para execução de experimentos laboratoriais com adultos de abelhas, que seja incluída na metodologia experimental o horário de montagem do experimento.

6 Referências Bibliográficas

BRIGHENTI C, R, G; CIRILLO A, M, BRIGHENTI D M. *Análise Longitudinal na determinação do fotoperíodo adequado para criação de abelhas em laboratório. Revista Bras Biom*, 26, 111-176 124, 2008.

IZBICK, R. & ESTEVES, L.G. *Análise de Dados Circulares*. In: IV Simpósio de Iniciação Científica e Pós-Graduação do IME-USP, 2008, São Paulo. Atas do IV Simpósio de Iniciação Científica e Pós Graduação, 2008.

LEAL, G.M.G. & SOUZA, F.A.M. Análise descritiva de dados de vento visando aplicações em plataformas. *Anais ...* In: 2º Congresso Brasileiro de P & D em Petróleo & Gás, 2003.

PRONI, E. A. & MACIEIRA, O.J. D. Ritmo circadiano da taxa respiratória de *Tetragonisca angustula fiebrigi* (Schwarz), *T. a. angustula* (Latreille) e *Trigona spinipes* (Fabricius) (Hymenoptera, Apidae, Meliponinae) *Revista Brasileira de Zoologia*, 21, 987-993, 2004.

OTIENO, B. S. *An alternative estimate of preferred direction for circular data*. Blacksburg, 2002. Dissertação submetida à Faculty of the Virginia Polytechnic Institute and State University.

R Development Core Team. R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, 2009.